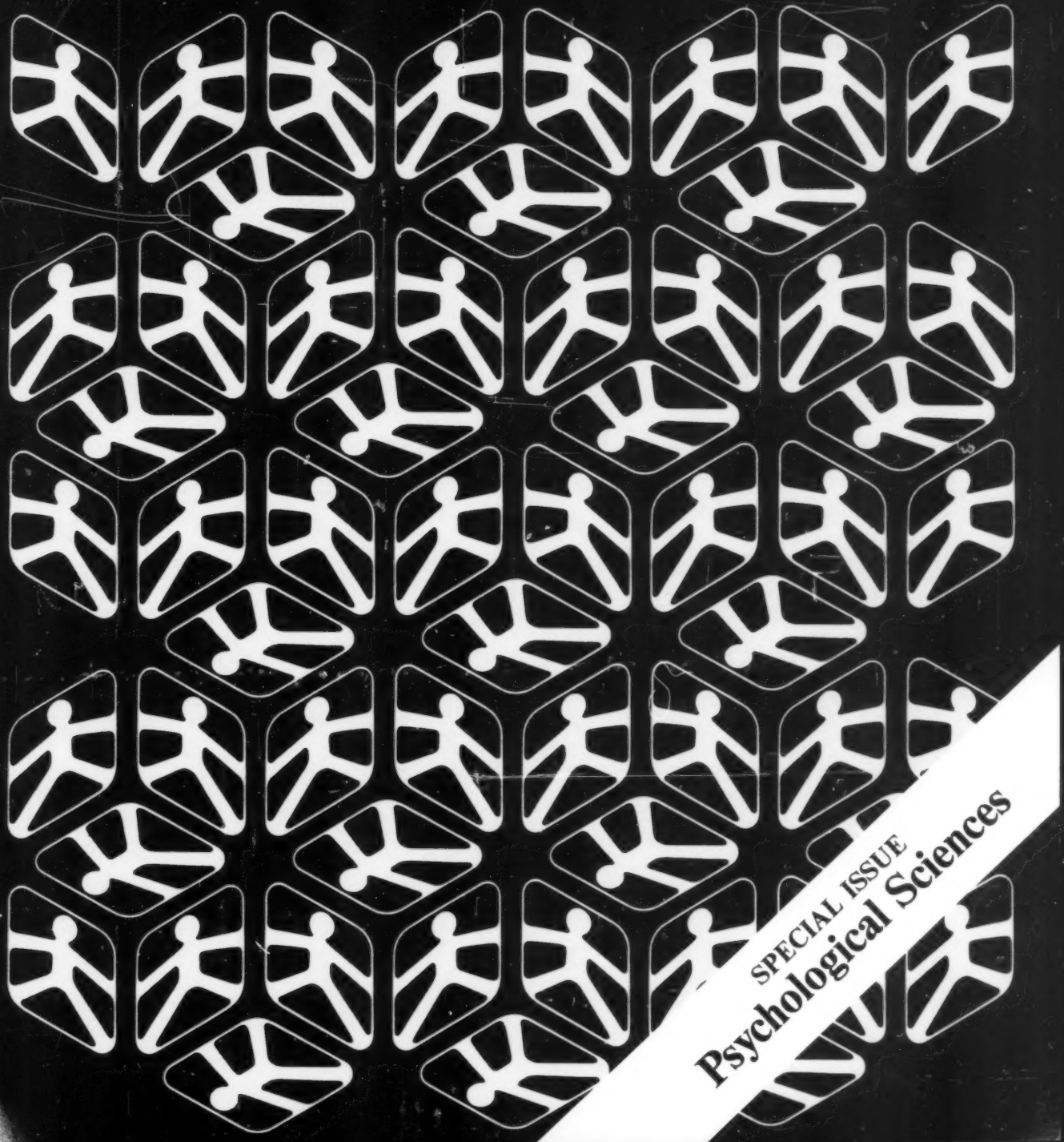
Naval Research Reviews

Office of Naval Research
One/1984
Vol XXXVI



SPECIAL ISSUE
Psychological Sciences

The undersea habitat SEALAB-II is being prepared for christening at the San Francisco Bay Naval Shipyard on 23 July 1965. SEALAB-II was the second phase of the Navy's Man-in-the-Sea-Program, which was organized and carried out by the Office of Naval Research (ONR). The SEALAB-II Project involved three ten-man teams living and working a total of 45 days at a depth of 205 feet one mile off the coast of La Jolla, California. During the summer of 1964, there was the SEALAB-I Project, which was much less ambitious in scope. From its early days and continuing through the present, ONR has supported research on the physiological and psychological effects of living and working in confined and unnatural conditions. Such research has enabled the Navy to enhance the productivity of personnel, as well as to improve their working conditions, as they go about such tasks as conducting deep sea salvage projects, studying marine life in its natural environment, and collecting geological and oceanographic data.

Naval Research Reviews

Office of Naval Research
One/1984
Vol XXXVI

Articles

Experimental Analysis of Team Performance by Henry H. Emurian, Joseph V. Brady, Ronald L. Ray, James L. Meyerhoff and Edward H. Mougey 3

The Psychology of Decision Making by Daniel Kahneman and Amos Tversky 20

Human Classification of Complex Sounds by James H. Howard, Jr. and John J. O'Hare 26

Cognitive Psychology by Marshall J. Farr 33

Intelligence and Mental Competence by Earl Hart 37

Departments

Introductory Note by James T. Lester, Office of Naval Research 2

Profiles in Science Amos Tversky 25

Research Notes 43

CHIEF OF NAVAL RESEARCH
RADM J. B. Mooney, Jr. USN

TECHNICAL DIRECTOR
Dr. M. K. Moss

CHIEF WRITER/EDITOR
W. J. Lescure

SCIENTIFIC EDITOR
Dr. G. A. Neece

ASSOCIATE SCIENTIFIC EDITOR
Dr. J. T. Lester

MANAGING EDITOR
M. J. Whetstone

ART DIRECTION
Typography & Design
Government Printing Office

ABOUT THE COVER

A multiple pattern design using the logo of the Human Resources Laboratory of The Johns Hopkins University School of Medicine. The lead article describes a research program on team performance supported by the Office of Naval Research at the laboratory. (Courtesy of The Johns Hopkins University.)

Naval Research Reviews publishes articles about research conducted by the laboratories and contractors of the Office of Naval Research and describes important naval experimental activities. Manuscripts submitted for publication, correspondence concerning prospective articles, and changes of address, should be directed to Code 732, Office of Naval Research, Arlington, Va. 22217. Requests for subscriptions should be directed to the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. *Naval Research Reviews* is published from appropriated funds by authority of the Office of Naval Research in accordance with Navy Publications and Printing Regulation NAVSO P-35.

Introductory Note

This issue is entirely concerned with work supported by the Office of Naval Research's (ONR's) Psychological Sciences Division. "Psychological Sciences" covers an enormous amount of territory, and ONR can fund only selectively. This introduction is designed to give an overview of what the Division is trying to accomplish, as background for the research reports that make up the rest of the issue.

Through many years' evolution, the Division has settled on three major sectors of psychology in which it is active. These are (1) Organizational Effectiveness, (2) Engineering Psychology, and (3) Personnel and Training Research. The names as well as the particular lines of inquiry receiving emphasis have been different at various points in ONR's history, but the same three sectors broadly defined have constituted the Psychological Sciences Program for some time. Investigators from each of these sectors are represented in this issue of *Naval Research Reviews*.

The Organizational Effectiveness Research Programs (once known as Group Psychology) are concerned with such topics as job performance, morale, retention, supervision, and quality of working life. They aim at an improved understanding of the fundamental variables that influence the important characteristics of groups of people working together. At the present time, this set of programs comprises 24 individual contracts with a total dollar value of some two million. These contracts are divided into clusters on the basis of central thematic ideas; the clusters change from time to time, depending on changing emphases. The present clusters are defined as Productivity, Effective Heterogeneous Groups, and Turnover and Rentention. The first article to follow, which is by Emurian, concerns team performance and is a sample of work funded by this sector of the Division.

The Engineering Psychology Programs are concerned with assuring effective performance by personnel operating and maintaining the high-technology equipment characteristic of modern Navy and Marine Corps systems. They seek better fundamental understanding of human perceptual and cognitive behavior in order to develop general guidelines for the design of compatible interfaces between people and their machines. At present these Programs comprise some 37 contracts, with a dollar value of some 3 million. These contracts are clustered as follows: Man-Machine System Interfaces, Visual and Auditory Perception, Information Processing and Decision Making, and Man-Machine Systems Technology. The articles by Kahneman and Tversky and by Howard and O'Hare are based on work funded by the Engineering Psychology sector.

The Personnel and Training Research Programs support research which will enable the naval services to more efficiently acquire, characterize, and utilize qualified personnel. Research efforts are focused on methods for assessing people's potential and competence, and for conducting rapid, efficient training. Two large considerations guide the efforts of this group. First, it appears that the best opportunities for increasing the efficiency of assessment and training lie in using the computer for tailoring each process to each individual. Second, the increasingly technological character of many Navy jobs suggests that cognitive and information-processing skills and aptitudes should receive major emphasis. These considerations have led the group to support research on objective, quantifiable theories of cognition and information-processing performances of the type required on Navy jobs. The program at present consists of 56 contracts with a dollar value of some 4.6 millions. The contracts fall into four main clusters: Theory-Based Personnel Assessment, Information-Processing Abilities, Instructional Methods and Technology, and Cognitive Processing. In this issue Marshall Farr, leader of the Personnel and Training Research Group at ONR, offers an orientation to cognitive psychology (which is the foundation for much of the Group's work), and Hunt provides an example of the work.

This small sample of the work of the Division can in no way convey the whole picture of its activities and efforts. We hope only that it may provide an increment of illumination about a segment of ONR's research programs which is, by reason of subject matter and methodology, often considered as not readily assimilable by much of the ONR community. The Division welcomes comments and inquiries concerning the work reported here as well as its other programs. An annotated listing of all current and recently completed contracts is available on request by writing to Code 442, Office of Naval Research, Arlington, Virginia 22217.

James T. Lester
Office of Naval Research

Experimental Analysis of Team Performance

by

Henry H. Emurian, Joseph V. Brady, and Ronald L. Ray
The Johns Hopkins University School of Medicine
and
James L. Meyerhoff and Edward H. Mougey
Walter Reed Army Institute of Research

Where the successful accomplishment of an organization's mission requires the coordinated contributions of two or more individuals collectively identified with the achievement of a common objective, the conditions for characterizing a team are operationally defined. While there has been much research on teams, for the most part observations of team performance under realistic or quasi-realistic conditions involving constant scenario environments have been limited by the constraints imposed on experimental interventions. Indeed, a review of the extensive literature in the area suggests that research on team performance effectiveness would benefit from the development and application of an effective methodology for extended-duration analyses of such situations, under conditions that provide for operational task assessment and evaluation within the context of a comprehensive living and work setting[1,2].

There is a growing recognition of the importance of developing technological guidelines related to (1) the impact of the type of mission, (2) the characteristics of team participants, and (3) the skill level of a novitiate participant. Of predominant importance are factors that affect a team's ability to accomplish mission objectives. Accordingly, a research project was undertaken to investigate performance effectiveness within the context of a laboratory environment in which both interpersonal and work behaviors can be continuously monitored and evaluated over extended time periods (*e.g.*, days). The project did not attempt to simulate a targeted operational environment. Rather, the laboratory facility was designed to address a broad range of performance problems from the perspective of a functional analysis of performance effectiveness. It is essentially a programmed environment with design features and measurement capabilities that permit the accurate assessment of relationships between antecedent conditions (*e.g.*, membership turnover, training methods, etc.), performance effectiveness, and consequences.

The conceptual framework within which the research was undertaken reflects the influence of three prominent classes of interacting factors: (1) *team composition*, to include personnel or membership characteristics (*e.g.*, number, gender, training, personal history, etc.); (2) *team resources*, to include facilities and physical setting factors (*e.g.*, hardware, living accommodations, communication networks, etc.); and (3) *team objectives*, to include performance programs and incentive conditions (*e.g.*, role assignments, pay-off matrices, etc.). These three broad categories of interacting factors are together representative of the range of theoretical and substantive issues addressed in previous team analyses and proposed for prospective team research agendas, with different investigations emphasizing one or another class.

This paper describes the experimental methodology and representative results derived from studies of such individual and team behavior. The studies were jointly sponsored by the Office of Naval Research and the National Aeronautics and Space Administration. The research methodology includes a laboratory environment that was intentionally designed to facilitate the implementation of a "behavioral program" of daily activities. The program not only structures the team participants' use of available resources but also provides the framework for the observation and measurement of a comprehensive range of behaviors.

Figure 1

The floor plan of the laboratory, and its position within the surrounding building shell.

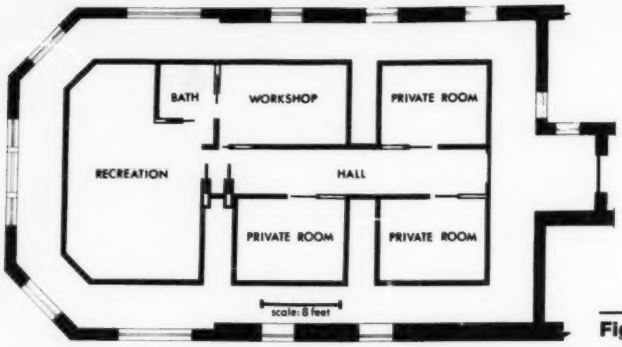


Figure 2

A participant's sleep area and desk along with storage drawers, intercom, cathode ray tube (CRT) communication console, entertainment console, etc.

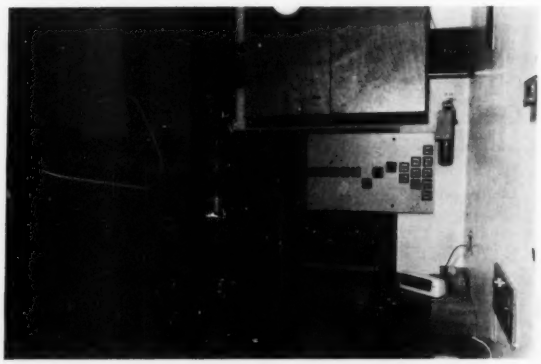
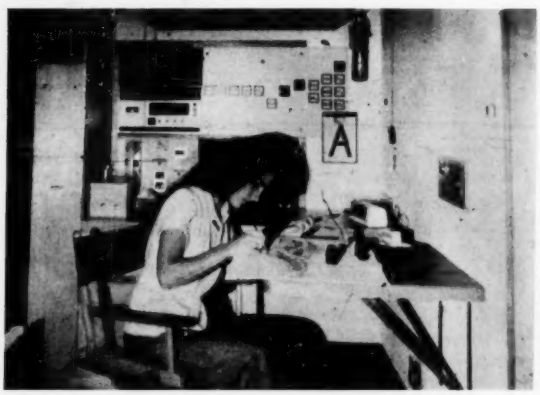


Figure 3

The opposite view of the same private room showing kitchen and bathroom-shower combination.

Figure 4

A representative participant at work on a creative manual task within a private room.



The Residential Laboratory Facility

The residential laboratory consists of five rooms and an interconnecting corridor, and it was constructed within a wing of The Henry Phipps Psychiatric Clinic at The Johns Hopkins University School of Medicine. The rooms are enclosed; they have their own walls and ceilings, but no windows, although access to an outside terrace can be granted as a research protocol may permit. The floor plan of the laboratory and its position within the surrounding building shell are presented in Figure 1. Each of the three identical private living quarters (2.6 × 3.4 × 2.4 m) is similar to a small efficiency apartment. Figure 2 shows a subject's sleep area and desk along with storage drawers, intercom, cathode ray tube (CRT) communication console, entertainment console, etc. The sleep surface, which can be enclosed by a privacy curtain, is normally covered so that the upholstery is exposed only during a sleep period. Figure 3 shows the opposite view of the same room. Shown are the kitchen and bathroom-shower combination. A door to the interconnecting corridor is to the left of this view, but on the right side of the room can be seen an unlocked full-sized exit door that leads to the perimeter of the laboratory. Figure 4 shows a representative study participant at work on a creative manual task in a private room. Figure 5 shows one view of the recreation area (4.3 × 6.7 × 2.7 m) that contains a complete kitchen facility along with exercise equipment and games. Figure 6 shows the opposite view of the same area that contains lounge chairs and an exit door to the perimeter. The workshop (2.6 × 4.1 × 2.7 m), presented in Figure 7, contains operator consoles for the individual and team performance tasks. A common bathroom serves the recreation and workshop areas. In summary, the programmed environment can accommodate at least three participants for intensive experimental analyses. More study subjects could be added to an experimental protocol by allowing them temporary residence within the recreation area along with periodic rotations to the privacy of the individual quarters when solitary members move to the recreation area.

The laboratory is "programmed" in the sense that its resources are restricted by design features that electronically regulate access to storage compartments or to areas containing supplies necessary to accomplish a given performance unit.

Figure 5

One view of the recreation room that contains a complete kitchen facility along with exercise equipment and games.

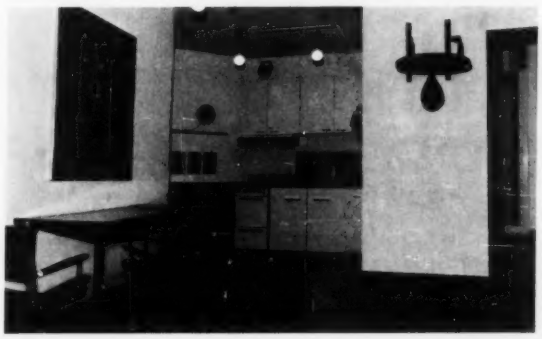


Figure 6

The opposite view of the recreation room showing lounge chairs and the exit door to the perimeter.

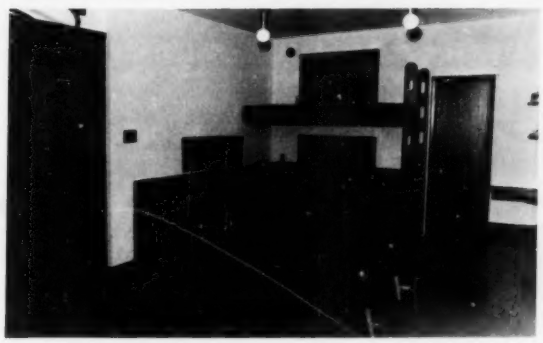


Figure 7

The workshop containing operator consoles for the individual and team performance tasks.

The Behavioral Program

To structure the team members' use of the laboratory's resources in a disciplined yet meaningful way, a behavioral program was developed to establish and maintain individual and team performance baselines as well as to provide the context for experimental manipulations of performance interactions during extended residential missions. A behavioral program is defined by (1) an array of activities or behavioral units and (2) the rules governing the relationships between these activities. Figure 8, for example, illustrates diagrammatically (1) the fixed and optional activity sequences that characterize a typical behavioral program used to establish baseline performances and (2) an array or inventory of component activities that constitutes such a program. Each box within the diagram represents a distinct behavioral unit and performance requirement, with progression through the various activities programmed sequentially from left to right. Finally, all behavioral units are scheduled on a "contingent" basis which means that access to a succeeding activity depends upon satisfaction of the requirements for the preceding unit.

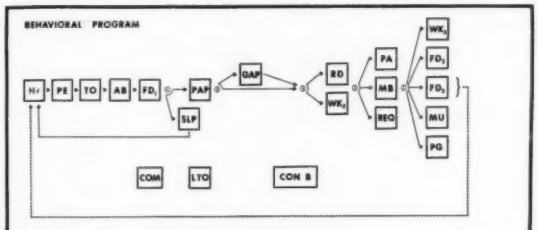
Beginning at the far left of the diagram, the fixed activity sequence is composed of all activities between and including Health Check (H✓) and Food One (FD1). The Health Check activity requires the subject to determine his temperature, pulse, and weight and to complete several status questionnaires regarding his mood and reactions to the laboratory environment. He then completes the following activities in the order displayed: Physical Exercise (PE), requiring 300 correct responses on an automated exercise task; Toilet Operations (TO), providing access to the private room bathroom and shower; Autogenic Behavior (AB), in which the subject follows taped relaxation instructions; and Food One (FD1), in which the subject is permitted to select two items from a presented list of 10 "light" foods such as coffee, soup, cereal, etc.

When Food One is completed, the subject is eligible to select *one* of two activities: Private Arithmetic Problems (PAP), requiring 150 solutions of problems presented on a CRT, or Sleep (SLP), providing access to the bed for an unlimited time period of at least 30 minutes. If the subject selects Sleep, he is required to return to the Health Check activity and the fixed activity sequence at the completion of Sleep. This is indicated by the broken line originating at Sleep and terminating at Health Check. In summary, then, the fixed activity sequence was designed to maintain and assess the subject's health if he were otherwise indisposed to engage in the broader selection of opportunities.

The optional activity sequence begins with the choice of Private Arithmetic Problems instead of Sleep. At the completion of that activity, the subject is eligible to select Group Arithmetic Problems (GAP), in which a subject can contribute to a performance criterion that must be satisfied before the group can enter the recreation area. Or the subject can skip Group Arithmetic Problems and choose *one* of the following two activities: Reading (RD), providing at least 30 minutes' access to books contained in a drawer, or Work Two (WK2), in which the subject completes, in private, various problems, experiments, or assembly projects presented in a drawer. When the selected activity is completed, the subject is eligible to select *one* of the following three activities: Puzzle Assembly (PA), requiring the subject to assemble a jigsaw puzzle presented in a drawer, Manual Behavior (MB), providing at least 30 minutes' access to art supplies contained in a drawer, or Requisition (REQ), allowing the subject to operate a manual task to earn points that are exchangeable for special privileges. On completion of the selected activity, the subject is eligible to select *one* of the following five activities: Work Three (WK3), providing access to the workshop, Food Two (FD2), requiring at least 30 minutes and providing the subject with a major meal to prepare and consume within his

Figure 8

A diagrammatic representation of the fixed and optional activity sequences that characterize a typical behavioral program used to establish baseline performances along with an array or inventory of component activities that constitutes such a program.



INVENTORY OF ACTIVITIES

NOTATION	FULL NAME	BRIEF DESCRIPTION
H✓	HEALTH CHECK	TEMPERATURE, PULSE, WEIGHT, STATUS REPORT
PE	PHYSICAL EXERCISE	300 CORRECT PRESSES ON AUTOMATED TASK
TO	TOILET OPERATIONS	USE OF PRIVATE BATHROOM AND CONTENTS OF DRAWER CONTAINING TOILETRIES
AB	AUTOGENIC BEHAVIOR	RELAXATION EXERCISES ON CASSETTE TAPE
FD1	FOOD ONE	TWO SELECTIONS FROM A LIST OF LIGHT FOODS
SLP	SLEEP	USE OF BED AND PRIVACY CURTAIN
PAP	PRIVATE ARITHMETIC PROBLEMS	150 CORRECT SOLUTIONS OF ARITHMETIC PROBLEMS
GAP	GROUP ARITHMETIC PROBLEMS	OPTIONAL, CONTRIBUTE CORRECT SOLUTIONS OF PROBLEMS TO GROUP RATIO CRITERION
RD	READING	ACCESS TO BOOK
WK2	WORK TWO	PROBLEMS, EXPERIMENTS, ASSEMBLY PROJECTS
PA	PUZZLE ASSEMBLY	ASSEMBLE A PUZZLE
MB	MANUAL BEHAVIOR	ACCESS TO ART MATERIALS
REQ	REQUISITION	EARN DELAYED DELIVERY OF TREATS
WK3	WORK THREE	ACCESS TO WORKSHOP
FD2	FOOD TWO	PRIVATE MAJOR MEAL
FD3	FOOD THREE	MAJOR MEAL IN RECREATION ROOM, GAMES
MU	MUSIC	EARN A CASSETTE TAPE
PG	PRIVATE GAMES	ACCESS TO SOLITARY GAMES
COM	COMMUNICATION	ACCESS TO INTERCOM
LTO	LIMITED TOILET OPERATIONS	ACCESS TO ESSENTIAL TOILET FACILITIES
CON B	CONDITION B	CHANGE IN PROGRAM CONDITION

private room, Food Three (FD3), providing at least 30 minutes in the recreation room by all subjects together, Music (MU), allowing the subject to earn a cassette tape that can be played at any time, or Private Games (PG), allowing at least 30 minutes' access to an assortment of solitary games within a drawer. As indicated by the broken line, once a subject completes his choice among those five activities, he returns to Health Check and resumes the fixed activity sequence. In summary, then, the optional activity sequence allows the subject flexibility in the selection and arrangement of activities, both individual and social.

At the bottom of the diagram there are two activities that have different selection rules. The Limited Toilet Operations (LTO) activity, which provides access to the bathroom but not the shower, can be selected at any time. The Communication (COM) activity allows access to the intercom for intersubject communications. A subject is permitted to use the intercom to initiate or answer a communication only if he is between activities within the behavioral program. Although the Communication activity is available between any two activities, a conversation requires at least two subjects' simultaneous presence within the Communication activity. Conversing subjects, however, can be located at different sequential positions within the behavioral program. For example, a Communication and conversation might occur when one subject is between Autogenic Behavior and Food One, and another subject is between Manual Behavior and the last column of activities.

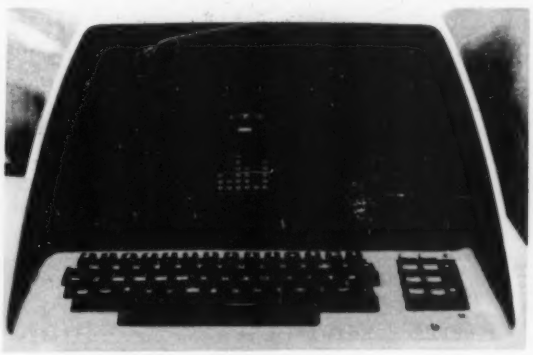
The behavioral program provides a promising approach to the problem of structuring the resources available to a confined microsociety. The functional interdependencies among activities ensure that performances of value to the welfare of the individual (*e.g.*, Physical Exercise), to the welfare of the team (*e.g.*, social recreation), and to the welfare of a "mission" (*e.g.*, sustained performance effectiveness) occur recurrently over time. These functional interdependencies reflect the "motivational" properties inherent within successive progressions through the program, and all incentives to maintain the overall operational status of the organization can reside within the behavioral schedule itself.

The behavioral program not only structures access to resources but also makes all corresponding activity units available for measurement. The boundaries between successive activities in the program impose rigor on the assessment of individual and group preferences and effectiveness within those activities. Additionally, the program has the advantage of providing a comprehensive range of variables for observation and measurement. For example, at one level, a subject's performance on arithmetic calculations could be assessed (*e.g.*, errors, response latency), and at another level, a subject's frequency and duration of progressions through the program could be assessed without regard to the intensive analysis of component activities composing such progressions. Moreover, the social status of the microsociety may be assessed by observing the degree of "synchrony" among subjects in the selection and completion of similar activities at the same time. Observations of subjects' communication networks along with the frequency, duration, and quality of dyadic and triadic social episodes would complement synchrony measures. All these factors, then, contribute to a method that has considerable and demonstrated power in the analysis of variables that impact upon individual and team performances, especially with regard to the potential interrelationships between the effectiveness of such performances and other aspects of the work environment.

Activities and performance requirements can easily be added to and withdrawn from the behavioral program. For example, it was desired to replace arithmetic calculations with a Multiple Task Performance Battery (MTPB) as a measure of complex human performance. As displayed in Figure 9, the MTPB is composed of the following five subtasks that are presented simultaneously to an individual

Figure 9

A photograph of the CRT console displaying the five subtasks composing the Multiple Task Performance Battery.



operator: (1) probability monitoring, (2) arithmetic operations, (3) warning light vigilance, (4) dynamic signal detection, and (5) target monitoring and recognition[3,4]. Accurate operation of the subtasks produces "accuracy points" that are cumulatively displayed on the CRT. To "build" this task into the behavioral program, a CRT console was located in the workshop in such a way that only one subject could perform the task at a time, and access to the workshop was made available between any two adjacent activities in the program. When a subject completed a work period in the workshop, he would return to his private quarters and resume the behavioral program at the point of departure. Private Arithmetic Problems and GAP were simply withdrawn from the inventory of activities.

Although performance effectiveness on the MTPB could be made a condition for access to other "high-value" activities in the program (*e.g.*, social recreation), a different incentive could be applied by relating it to a subject's compensation. The latter approach has been used for recent investigations in the laboratory. Thus, the intrinsic motivational properties of the behavioral program provide the context in which external incentives can be applied to obtain a direct moment-to-moment relationship between performance effectiveness (*e.g.*, quality and quantity) and its immediate consequence. Such an interplay between incentives can be dramatically effective in (1) generating and sustaining complex human performances over extended time periods and in (2) providing the ancillary contextual observations that make performance changes interpretable in terms of a functional analysis.

Subjects respond favorably to the programmed environment and to the behavioral program. With few exceptions, they report a sense of heightened accomplishment and productivity in the use of their time while following the program. Although there are clocks present in each room of the laboratory, the behavioral program is not oriented to specific time markers, and subjects may drift in wake-sleep cycles according to their personal dispositions. If sleep discipline is required, however, the program could anchor the Sleep activity to a particular time window (*e.g.*, 2300–0600 hours), and remaining time would be filled with other programmed activities. Such a procedure has been used successfully in an investigation that required subjects' circadian rhythms to be held constant[5].

Participants

Well over one hundred male and female volunteers have participated in the research program to date, and only three subjects have withdrawn from an experiment before its scheduled completion. Almost all participants have a college background, and many are graduates. Acceptance into the research program follows psychological evaluation and detailed orientations to the laboratory and to the behavioral program. The research involves no elements of deception, and informed consent is an integral component of the orientation. Unless otherwise noted, the research results to be summarized herein were based upon analyses of three-man groups.

Social Factor Effects

Having demonstrated the utility and acceptability of the behavioral program in a series of six-day to sixteen-day residential investigations with two and three subjects, we directed our attention to ten-day analyses of the importance of social factors as they affect the status of a three-person microsociety. These analyses examined social contingencies that required coordination among group members before access to the recreation area was granted. The results showed that such contingencies, embedded within the behavioral program, could counteract a team's tendency to fragment over time[6]. The contingencies prevented persons who had little interest in interacting socially from becoming isolated from the group and, in some cases, from showing a decline in individual performance effectiveness[7]. Related experiments showed the greater reinforcing strength of triadic social episodes over dyadic episodes, and they also indicated a relationship between social distance in a triad and time spent in a dyadic situation when social opportunities were limited to pairs of subjects.[8]

Taken together, these studies indicated that low group cohesiveness increased members' vulnerability to social fragmentation in the absence of specifically programmed triadic contingencies of reinforcement designed to promote productive social interactions among team participants.

Incentive Effects

The preceding investigations were undertaken with incentives inherent within the behavioral program (subjects received a per diem allowance). In the next set of studies, the interplay between incentives, both internal and external to the program, was introduced as the means of sustaining individual and group behavior. These studies were designed to develop a laboratory model that would allow systematic exploration of individual and social by-products of avoidance incentive conditions. Under a positive incentive condition, "work units" (*e.g.*, physical exercise, manual operations, MTPB performance) were completed by individual team members, and each such unit added a fixed amount to a group account that was divided evenly among the three subjects at the conclusion of the experiment. Under an avoidance incentive condition, however, no money was earned, mission members were assigned a daily performance criterion to satisfy as a team, and failure to reach the criterion resulted in reductions in accumulated earnings. The two incentive conditions appeared in various orders and durations across the series of investigations.

Comparisons between conditions on a number of behavioral program measures dramatized the deleterious effects of the avoidance incentive condition. Disruptive by-products of that condition included (1) interpersonal confrontation and antagonism, especially by high-productivity subjects toward low-productivity subjects, (2) vociferous written and spoken complaints about the schedule, (3) written and spoken hostility directed toward the experimenters, and (4) dysphoric feelings[9]. The impact on performance effectiveness during the avoidance condition was demonstrated by one team member who refused to work further on the MTPB when another member fell somewhat behind in his share of work as agreed upon by team participants[10]. In contrast, under positive incentive conditions, such disruptive effects did not occur even when extraordinary performance productivity was observed, and a several-day history of negative effects could be overcome by reintroducing the positive condition. These effects emphasized the interaction between heterogeneity in work productivity within an organization and member tolerance and intolerance of such heterogeneity under different incentive conditions.

The foregoing investigations clearly established social variables as fundamental contributors to the overall status of a confined microsociety, and they emphasized the sensitivity of such variables to a range of experimental manipulations having significance for real life operational conditions. Throughout such studies, it was observed that mission participants sought social interaction under one set of conditions (*e.g.*, triadic social contingencies and positive performance outcomes) and withdrew from such interaction under other conditions (*e.g.*, pairing social contingencies and avoidance performance outcomes). Thus, the joining and leaving of a group by mission participants, under circumstances encompassing more than a single environmental condition, appeared to generate social effects that reflected important dynamic processes requiring systematic experimental analysis.

Extra Person Introduction/Withdrawal

Accordingly, team performance effectiveness studies were initiated to assess the effects on individual and group behavior of a novitiate participant's introduction into and subsequent withdrawal from a previously established and stable two-person social system[5]. The objectives were to focus upon (1) the social mechanisms and temporal properties associated with the integration of such a participant into an established team and (2) sources of group disruption or cohesiveness fostered by his or her presence. Additionally, we began to measure changes in level of urinary testosterone as an endocrinological index of demonstrated sensitivity to social interaction effects in both animals[11,12] and humans[13,14]. This behavioral-biological analysis was used to provide a more comprehensive assessment of the personal and social impact generated by the introduction and withdrawal of new members with an established group[15].

Urinary testosterone levels were determined by radioimmunoassay. Following a 72-hour hydrolysis with beta glucuronidase, the samples were extracted with methylene chloride. The methylene chloride layer was washed with water and dilute sodium hydroxide and then evaporated. The extracts were purified on LH-20 Sephadex columns. Recoveries through the procedure were monitored by the addition of a small amount of tritiated testosterone added to each sample prior to extraction. The Sephadex column eluates were evaporated and taken up in RIA buffer. Aliquots were incubated overnight at 4°C with a testosterone antibody produced in rabbits. Free and antibody-bound hormone was separated using Somogyi reagents. Radioactivity measurements were made in a Beckman LS-20 counter. Samples were assayed in duplicate and corrected for recovery.

The paradigm adopted for experimental analyses of effects of changes in group size and composition was as follows. A two-person group resided for ten successive days within the programmed environment, and the members operated performance tasks for their earnings. During that ten-day period, a third "novitiate" participant was introduced into the programmed environment for several successive days, increasing the group to three members. A typical "introduction" period with three group members lasted four days, and it usually began on Day 4 or Day 7 of a ten-day experiment.

The rule conditions associated with the novitiate's entrance into the group differed across successive investigations. In some studies, the novitiate received a per diem allowance and was not required to work for compensation, although he was permitted to contribute to the performance tasks that benefited the two established group members. In other studies, the novitiate was required to work for compensation by competing with the two other group members for access to the single MTPB console in the workshop. Finally, there was a series of investigations with both male and female novitiates and, in some cases, with novitiates and dyadic members who had previously participated in a residential study.

In studies where the novitiate's presence served primarily as additional social stimulation for the established dyad and as a source of information regarding current events outside the laboratory, the two-person group initially showed a resistance to granting the novitiate permission to work, even when such work would have provided relief from operating a demanding task. However, as the three-person condition continued over days, novitiates were observed to contribute to work productivity to a degree that was almost equivalent to the productivity of the dyadic members. Since there were no external incentives for a novitiate's work in these first introduction studies, these findings emphasized the influence of social processes alone in maintaining performance productivity, at least within

these cohesive group situations. Finally, novitiates showed daily urinary testosterone at the upper and lower boundaries of the standard range, but the absence of baseline levels precluded the interpretation that active social processes had governed such effects.

Transitions between two-person and three-person conditions were not always smooth in groups where the novitiate had to work the MTPB for compensation. When a novitiate forcefully intruded himself into the dyad's customary work schedule, his testosterone levels rose or fell generally in close relationship with his success or failure, respectively, in gaining and maintaining access to the MTPB station according to a schedule that was least disruptive to his wake-sleep cycles, as determined during several baseline days preceding his introduction into the group. When sleep discipline was imposed, and when a novitiate was cooperative in negotiating an orderly sequence of using the MTPB, there were no notable changes in testosterone observed in any team participant. Finally, when a female novitiate was introduced into a two-man group, wake-sleep cycles and work periods were erratic throughout the three-person condition. Such effects were associated with the absence of notable androgen changes, even by a dyadic member who, as a novitiate in an earlier study, had successfully maintained his wake-sleep cycles and had shown a striking increase in testosterone when he joined the group.

From such an account, principles emerge that can facilitate the development of technological guidelines designed to minimize the disruptive effects of transitions in group size and organization.

Group Member Replacement

The next series of experiments demonstrated the extension of the research paradigm from analyses of "introduction" effects to the analysis of "replacement" effects. That is, the group size was held constant to evaluate effects of *replacing* a member of an established three-person group with a novitiate participant. These replacement analyses were systematic replications of the earlier studies. In a research strategy based upon systematic replications, as compared with exact or direct replications, effects of the experimental variable or treatment are demonstrated by affirming the consequent[16], in which case each successive replication incrementally contributes to an understanding of effects that can be reliably attributed to the antecedent condition (*e.g.*, introductions or replacements). The generality of the behavioral processes is assured by showing similar relationships across a broad range of circumstances (*e.g.*, subjects, order and duration of experimental conditions, performance tasks, group size, etc.). This research strategy, as adopted by the programmed environment unit, has proved to be most productive and economical, an important consideration in light of the expense and staffing effort required to undertake programmed environment investigations.

A typical replacement investigation proceeded as follows. A three-person group resided in the programmed environment for five successive days. At the end of Day 5, one person was withdrawn and replaced by a novitiate participant who, with the remaining two original members, formed a new group for the next five successive days. Consecutive studies differed in terms of (1) the decision rule by which an original group member was withdrawn, (2) the number of baseline days that came before group formation, and (3) the type of performance tasks that the group members operated for compensation.

First Replacement Experiment

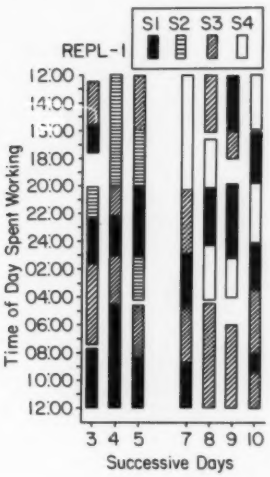
For the first replacement experiment (REPL 1), three-person group members resided in their private rooms for a two-day baseline "alone" period. During this time, access to the intercom, to social activities, and to the MTPB work station was prohibited. This two-day period provided a necessary hormonal reference against which to assess endocrine responses in relationship to initial group formation. On Day 3, all previously prohibited activities were made available to the group, and each member was required to operate the MTPB for individual compensation. As in the introduction experiments, there was only one MTPB console located within the workshop, and subjects occupied the workshop singly on a self-determined rotational basis. This procedure, then, permitted an evaluation of the manner in which subjects occupied the work station (*e.g.*, duration of work periods, time-of-day of work periods, etc.) as one of the principal dependent variables of the experiment.

At the end of Day 5, the mission member who had earned the fewest MTPB performance points, totalled across Days 3–5, was withdrawn from the experiment. This decision rule was known by the group members before the experiment began. The novitiate participant entered the programmed environment on Day 6, which was a solitary baseline day for all three subjects. On Day 7, the newly formed team had access to intercom communications, social activities, and the MTPB work station that continued to be available throughout Days 7–10. Thus, the two ten-day participants were required to adjust to the replacement of an original member, and the novitiate member was required to adjust to his entrance into an established unit whose members shared a history of having competed successfully to maintain high levels of performance effectiveness.

Figure 10 presents time of day spent working on the MTPB for all subjects across successive days of the experiment when access to work was permitted. The novitiate participant is identified as "S4." Throughout Days 3–5, subjects alternated in their occupancy of the work station, with uninterrupted work periods ranging from 2 hours (*e.g.*, S1, Day 3) to 9 hours (*e.g.*, S2, Day 4). The lengthy work period exhibited by S2 on Day 4 was related to his

Figure 10

Time of day spent working on the individual Multiple Task Performance Battery for all subjects across successive days of the experiment (REPL 1) when access to work was permitted.



attempt to remain competitive after having worked only 2 hours on Day 3. When the novitiate (S4) began to work on Day 7, having replaced S2, he initially preempted the work station for at least nine uninterrupted hours of MTPB performance. Other group members were unappreciative of this intrusion and revealed this by negative interpersonal ratings assigned to S4 during the Health Check activity. Thereafter, the novitiate and the remaining group members alternated occupancy of the work station, with S3's work times clearly later in the day than they were during Days 3–5. Neither the original group nor the re-formed group showed stability across days of work times. This outcome is perhaps attributable to the competitive contingencies for individual compensation that were present throughout all work days.

Figure 11 shows time of day spent sleeping for all subjects during the experiment. Comparatively stable sleep patterns were exhibited only by S2 who showed uninterrupted sleep episodes beginning between 2400 and 0500 hours across Days 1–5. During the same five-day period, Subjects 1 and 3 almost always showed erratic sleep episodes that differed across days in time of day of occurrence, frequency, and duration. Similar erratic patterns persisted during Days 6–10 when S2 was replaced by the novitiate (S4).

Figure 11

Time of day spent sleeping for all subjects across successive days of the experiment (REPL 1). Bracketed days [1], [2], and [6] were baseline "alone" days.

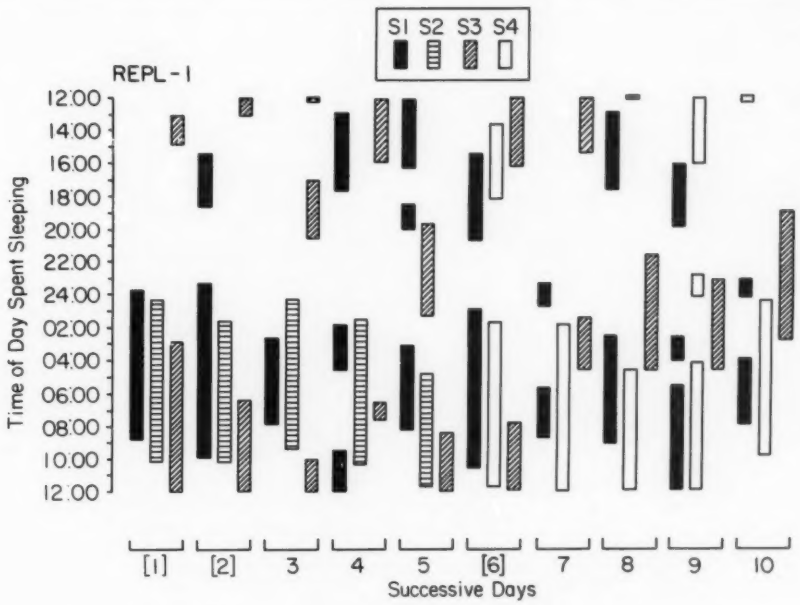


Figure 12

Total urinary testosterone for all subjects across successive days of the experiment (REPL 1).

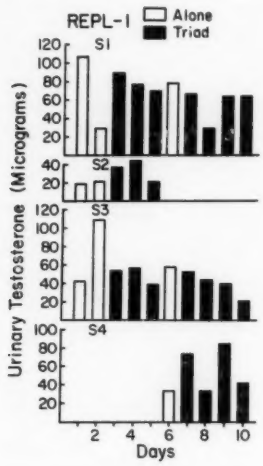
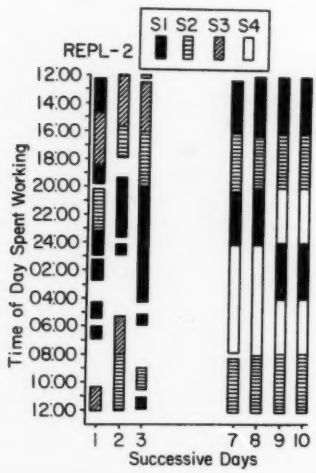


Figure 13

Time of day spent working on the individual Multiple Task Performance Battery for all subjects across successive days of the experiment (REPL 2) when access to work was permitted.



Importantly, the novitiate showed a clear reorientation in his sleep episodes that persisted throughout Days 7–10. These latter effects reflect the readjustments that were required by at least one original group member when the novitiate became a working participant during Days 7–10 of the experiment.

Figure 12 shows total urinary testosterone for all four subjects. With respect to the original group, S2 showed testosterone values that were somewhat lower than the other two participants. Importantly, these comparatively lower values were evident during the first two baseline days. When group members commenced working on Day 3, S2's values increased somewhat over baseline levels, but they continued to be lower than the values exhibited by the other two members across Days 3–5. Significantly, S2 was the mission member who did not compete successfully to remain within the experiment for ten days, and he was withdrawn at the conclusion of Day 5. Finally, across Days 7–10, testosterone levels progressively declined for S3 in relationship to his shift in work and sleep times. This latter effect confirms the outcomes observed in the introduction studies, and it demonstrates, by systematic replication, the generality of the behavioral-biological processes governing such effects.

Second Replacement Experiment

The experimental design of the second replacement analysis (REPL 2) was similar to the first with two major differences. First, the novitiate group member was a female who had previously participated in an unrelated ten-day residential experiment, and she had almost 60 hours' practice on the MTPB. Second, to provide more days for competition to remain in the experiment and a longer history of sustained performance effectiveness by two group members prior to the novitiate's entrance, no initial baseline was programmed. The novitiate, then, entered the environment at the beginning of Day 6, which was a baseline day for all subjects, with more experience in the laboratory than the two other group members. Thus, the two ten-day participants were required to adjust to the replacement of an original group member by a person having extensive programmed environment experience.

Figure 13 presents time of day spent working on the MTPB for all subjects on days when access to work was permitted. The novitiate participant is identified as "S4." Throughout Days 1–3, subjects alternated occupancy of the work station in an erratic fashion within and across days, with work periods lasting from 1 hour (*e.g.*, S1, Day 1) to 8 hours (*e.g.*, S1, Day 3). Subject 3 voluntarily withdrew from the experiment during Day 3, reasoning that his performance would not result in his participation beyond Day 5. Since the novitiate was not scheduled to appear until Day 6, a baseline day for all subjects, the two remaining subjects were programmed with baseline days on Days 4 and 5. This

preserved the integrity of the experimental design plan in relationship to analyses of three-person working groups. In striking contrast to work times during Days 1–3, work times during Days 7–10 were orderly and precise. The pattern for Day 8 is identical to Day 7, and the pattern for Day 10 is identical to Day 9. Throughout Days 7–10, all subjects occupied the work station for eight hours each day.

These data show the impact of an experienced person, who exhibited assertiveness and leadership, on an established group whose members had previously competed successfully to remain within the experiment. Although the two-person group followed the suggestions, if not the directions, of the novitiate, S4 received negative interpersonal ratings on the Health Check questionnaires.

Figure 14 presents time of day spent sleeping for all subjects. Although sleep times were perhaps not as erratic as those in the previous experiment, only S2 showed patterns that were somewhat consistent across all mission days. Additionally, the novitiate shifted her sleep pattern on Day 8, and thereafter commenced sleep periods in the early hours (*e.g.*, 1200) of an experimental "day." Finally, the stable sleep patterns exhibited by all subjects on Days 9 and 10 corresponded to stable work periods also observed on those two final mission days.

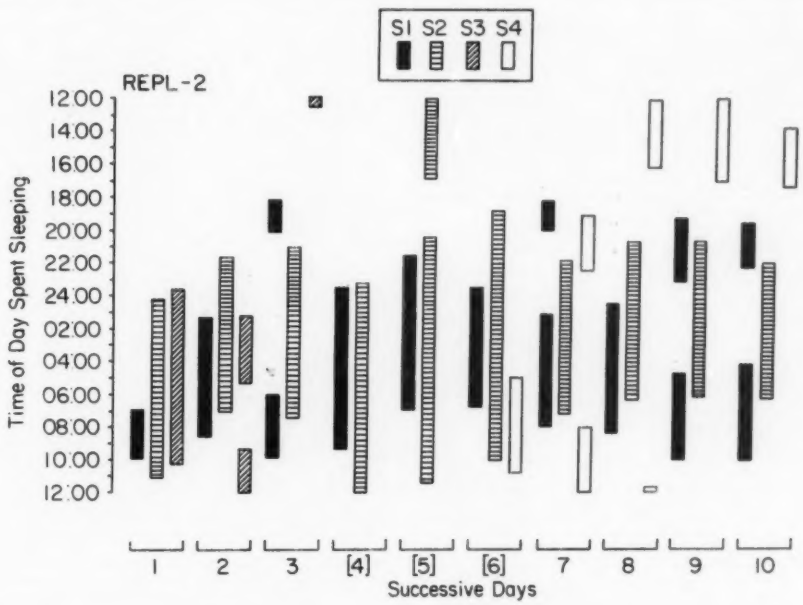
Third Replacement Experiment

In all previous investigations, the coordination required of mission participants was reflected in the sequential use of the work station and in the program synchrony necessary for subjects to meet together in the recreation room. In the next replacement investigation (REPL 3), however, a team performance task was introduced into the research protocol that systematically replicated the preceding analyses with a task that involved far more stringent coordination requirements.

The team performance task is an expanded version of the single-operator MTPB that previously served as the project's principal performance assessment tool. The Team MTPB (TMTPB) involves three operator consoles, each of which presents the identical display of the five task components (see Figures 7 and 9). The parameters of these tasks were modified to a difficulty level such that the concurrent inputs of three operators were required to avoid information overload and to produce maximum performance effectiveness per unit time. The "team" aspect of the task is reflected by the interlocking response demands associated with the probability monitoring subtask, and it is embedded within the context of the remaining four individually solvable subtasks. The team subtask requires the detection of a

Figure 14

Time of day spent sleeping for all subjects across successive days of the experiment (REPL 2). Bracketed days [4], [5], and [6] were baseline "alone" days.



bias that is recurrently presented on any one or more of the four probability monitoring scales. To "correct" a bias, each of the three operators must press the corresponding "correct" keyboard character within 0.6 second of the first such keyboard entry. Correction of a bias produces increments in accuracy points, and a team's failure to detect and correct a bias results in subtractions from accumulated points. The team task, then, requires (1) processing of symbolic information (*i.e.*, the detection of a bias), (2) sharing information by communications among team members (*e.g.*, One operator may say "Bias on one. Ready. . .Go."), (3) coordination of a response (*i.e.*, three response inputs within 0.6 second), and (4) sustained vigilance to avoid loss. This team task reflects the major performance dimensions considered to be crucial to developing methods for quantitative analyses of the interrelationships between individual and team performance effectiveness[17].

The ten-day experiment began with a three-man team whose members were new to the programmed environment and to the TMTPB. Participants had been acquainted with the *individual* MTPB during an orientation session, but acquisition of the TMTPB occurred for the first time on Day 1 of the experiment. For remuneration for participation, the team operated the TMTPB to a performance ceiling of 5000 accuracy points each day, requiring 6–9 hours of work to accomplish. The team members decided among themselves the manner of distributing the performance demands of the individual and team subtasks.

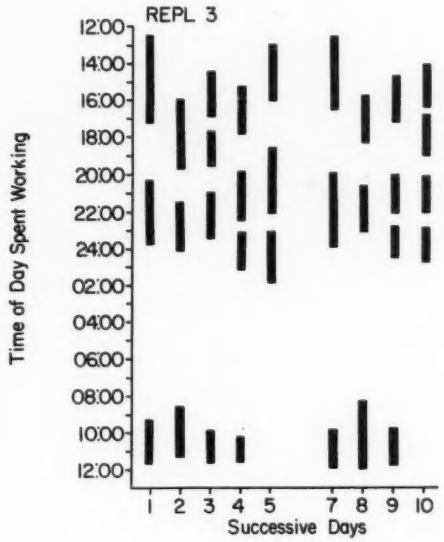
At the end of Day 5, one of the three original team members was withdrawn from the experiment. Initial team members began the study with the understanding that one participant would be withdrawn, but they were not given the decision rule by which that choice would be made. At the beginning of Day 6, then, a novitiate participant was introduced into the programmed environment. To accommodate this transition, the three participants followed the behavioral program in their private quarters on Day 6, but without access to the TMTPB, intercom communications, and social activities. On Day 7, the novitiate member joined the team as the replacement participant, and this newly formed team operated the TMTPB on Days 7–10.

Figure 15 presents time of day spent working by the team. This figure shows that three or four work periods occurred each day, and they ranged in duration from two to five hours. Although the time of day associated with work periods differed across days, work was not generally observed between 2400 and 0800 hours of a day. Finally, the pattern of work that the initial team adopted was also observed during the final four days of the study with the reformed team.

Figure 16 presents time of day spent sleeping for all subjects. The novitiate participant is again identified as "S4." Although the behavioral program was not oriented to

Figure 15

Time of day spent working on the Team Multiple Task Performance Battery across successive days of the experiment (REPL 3).



time markers, sleep periods were generally stable across successive days for both original and re-formed teams. When drift in sleep onset time occurred across days, all members of a team drifted in concert with each other.

The dynamics of the components of the individual and team subtasks differed. Figure 17 shows, for example, points earned on the individual subtasks of the TMTPB across successive work periods. This figure graphically shows smooth initial acquisition (Segment 1) and reacquisition (Segment 2) trends on the individually solvable subtasks. It also shows that the re-formed team exhibited degraded performance during the first two work periods of Segment 2 and that performance reacquisition was more rapid than was acquisition by the original team. Performance on the individual subtasks was degraded despite the presence of two team members who had a combined total of almost eighty hours' practice on the TMTPB. That such performance degradation was not associated with social disruption was indicated by the absence of negative ratings toward the novitiate as determined from recurrent Health Check assessments.

Figure 16

Time of day spent sleeping for all subjects across successive days of the experiment (REPL 3). Bracketed day [6] was a baseline "alone" day.

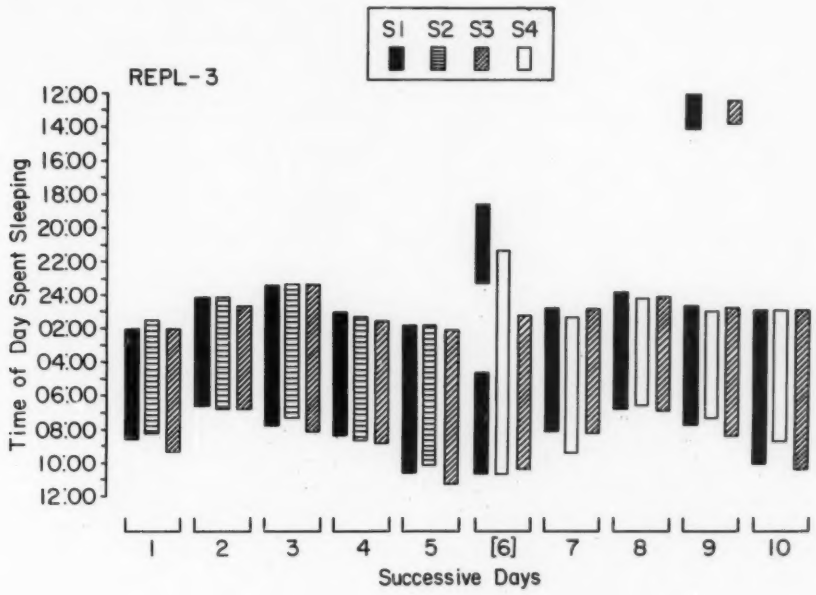


Figure 17

Points earned on the four individual subtasks of the Team Multiple Task Performance Battery across successive work periods.

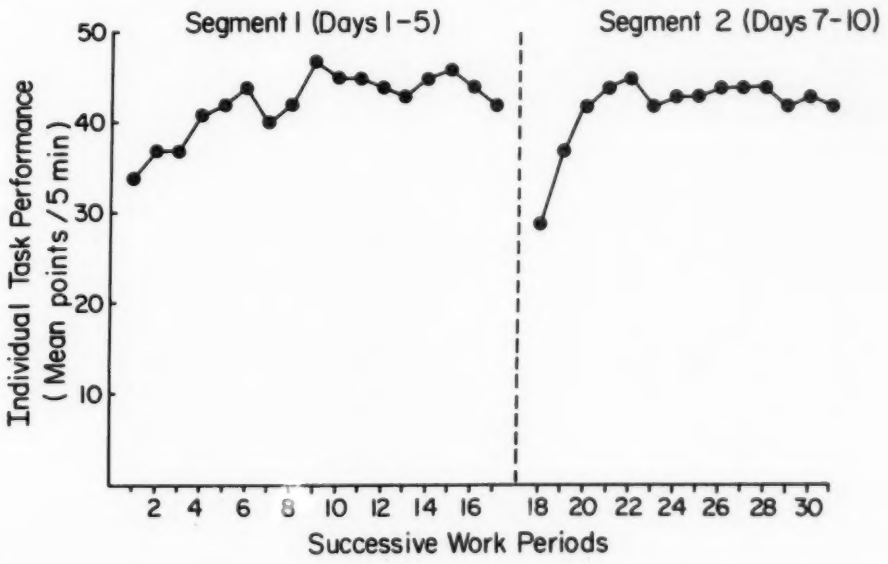


Figure 18 shows points earned on the team subtask of the TMTPB across successive work periods. In contrast to individual task performance, performance effectiveness on the team task was erratic, even though a trend toward improved team performance is graphically apparent for initial (Segment 1) and re-formed (Segment 2) teams.

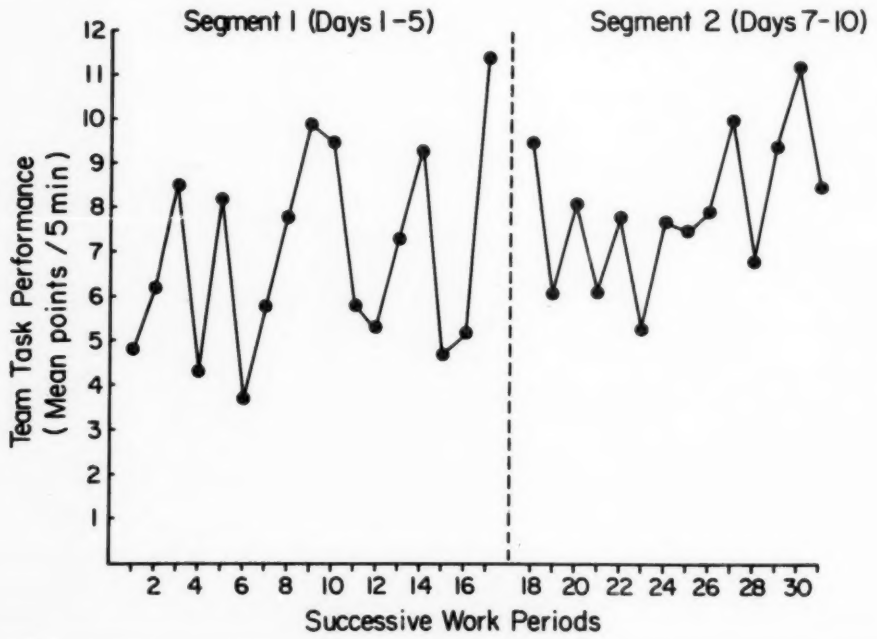
These observations suggest that improvement in *combined* individual and team performance effectiveness over successive work periods was attributable, in large part, to improvement on the contextual individual subtasks. Additionally, preliminary analysis shows that improvement on the team subtask for the original team (Segment 1) was attributable to a progressive "sharpening" of discrimination in terms of fewer false alarms over successive work periods. Such was not the case, however, for the re-formed team (Segment 2). During Days 7–10, whatever improvement there was on the team subtask was attributable to fewer missed biases and not to fewer false alarms. Thus, a clear shift occurred in the operation of the team subtask between Segments 1 and 2, despite the overall trend toward improved performance across both segments.

The performance shift observed on the team subtask between Segments 1 and 2 suggests the involvement of a more complex process of acquisition and reacquisition than repeated practice. Such a process might involve "solution strategy rehearsals" among team members for responsibility in operating the several subtasks. For example, when the novitiate joined the team, the two original team members probably assumed rotational responsibility for monitoring the team subtask to avoid potential losses while the novitiate mastered the discrimination. Finally, the prominent involvement of such rehearsals and rotations is further indicated by the fact that progressive improvement in overall performance effectiveness was attributable, for the most part, to improvement *between* successive work periods rather than to improvement *within* work periods.

Much more needs to be learned about those strategies and rotations and their dynamic interplay with individual and team performance effectiveness. Against the background of the introduction analyses that showed an established team's resistance to accepting a novitiate's work, the present study shows that a novitiate's lack of skill on a task can perhaps be masked by experienced team members who are unwilling to tolerate even a temporary degradation in

Figure 18

Points earned on the team subtask of the Team Multiple Task Performance Battery across successive work periods.



overall team performance effectiveness. The penalty of such a strategy lies in the constraints on redundancy of skills that could result in even more drastically degraded performances under conditions of further replacements of the original team members. With quantitative (*i.e.*, computer assisted) means of assessing the moment-to-moment performances of team members, the relative contributions of individual member performance to the terminal steady-state of the system could be characterized. At the very least, such a characterization would suggest intervention guidelines or pre-training schedules that would ensure the most effective balance between individual and team performance effectiveness and subtask proficiency under the various conditions of membership turnover.

Conclusions

The objectives of this research project on individual and team performance effectiveness have been the development of principles and procedures relevant to the selection and training of mission personnel. The research findings were based upon the investigation of preventive monitoring and corrective procedures to enhance mission performance effectiveness and upon the evaluation of countermeasures to the potentially disruptive effects of high-demand and stressful performance, interpersonal, and physical environments. Initial research activities were directed toward the design and development of an experimental microsociety environment for continuous residence by three-person groups of human volunteers over extended time periods. The environmental conditions provided for programmable performance and recreational opportunities within the context of a biologically and behaviorally supportive setting. Studies were then undertaken to analyze experimentally (1) conditions that sustain group cohesion and productivity and that prevent social fragmentation and individual performance deterioration, (2) motivational effects produced by the programmed incentives maintaining individual and team performance requirements, and (3) behavioral and biological effects resulting from changes in team size and composition. The significance of these investigative endeavors is to be understood in terms of emergent motivational, social-interaction, and group composition principles having practical relevance to the establishment and maintenance of operational mission performance effectiveness.

In this latter regard, the results obtained from these small-group studies clearly established the applicability and generality of behavioral technologies and methodologies to the experimental analysis of individual and team performances within the context of a human microsociety. Additionally, behavioral programming techniques were demonstrably effective in generating and maintaining such individual and group performances for unobtrusive monitoring and measurement with precision and regularity over time. Furthermore, the interplay between incentives both internal and external to the program provided the occasion for observations of performance in relationship to realistic incentive schedules. Contingency management principles, along with the technological guidelines that provided the basis for design and development of the programmed microsociety environment, were shown to be capable of sustaining individual and team performance effectiveness and group cohesion without notable biological or behavioral disruption under conditions of spatial restriction, social separation, enforced intimacy, and high performance requirements.

More specifically, the results of these studies showed that both individual and group productivity can be enhanced under confined microsociety conditions by contingency management principles applied to designated "high-value" component tasks embedded within the overall performance program. Similarly, group cohesiveness can be promoted, and individual social isolation and/or alienation (*i.e.*, group fragmentation) can be prevented by the application of contingency management principles to social-interaction segments of the performance program.

Conditions that were found to result in progressive deterioration of individual and team performance effectiveness included aversive programming contingencies such as loss of accumulated earnings. The by-products of aversive schedules that emerged under such circumstances were found to be detectable and quantifiable in measures of verbal performance (*e.g.*, behavioral program ratings), interpersonal performance (*e.g.*, verbal confrontation and aggression), work performance (*e.g.*, diminished productivity), and team morale (*e.g.*, irritability and dysphoric mood). In contrast, positive incentive schedules effectively counteracted the disruptive consequences of aversive contingencies while at the same time supporting high work productivity free from negative side-effects.

Related research results emphasized the prominent involvement of behavioral and biological processes that were functionally related to adjustments and reactions when changes occurred in group size and team composition. The experimental analysis of such "introduction" and "replacement" effects emphasized the critical importance of providing a structured transition in the form of orientation and training regimens for both novitiate and established team participants to minimize potentially disruptive effects of altering the interpersonal and social dynamics of a microsociety.

The development of the Team Multiple Task Performance Battery (TMTPB) opened an important dimension to the research program. By imposing task requirements involving coordinated responses among team members, the analysis of the dynamic interplay between individual and team performance effectiveness was initiated within the context of ongoing "replacement" investigations. For example, it was found that the replacement of an established team member by a novitiate participant resulted in degraded

performance on the individual subtasks of the TMTPB and in a shift in the operation of the team subtask itself. These findings suggest that a novitiate's lack of experience and skill on critical aspects of a task requiring coordination could be masked by experienced team participants who were perhaps unwilling to tolerate even a temporary degradation in overall team performance effectiveness. By developing methods to detect and quantify the relative contributions of individual members to the operation of the team task, intervention guidelines or pre-training schedules could be investigated that would ensure the most effective balance between individual and team performance effectiveness and subtask competency under the various conditions of membership turnover.

Acknowledgements

Operational aspects of the research laboratory are coordinated and administered by Ann Rutledge. Jerry Locklee is the principal supervisor of ongoing investigations, and June Hitt assists with the many details associated with a study's successful completion. Electrical engineering and computer sciences input is provided by Dave Krausman along with Rick Wurster, Ron Atkinson, and Leonard Daley. The following individuals have assisted in monitoring the conduct of experiments during the past two years: Richard Bodek, Anne Campbell, Kathleen Connors, Osbert Cush, Timothy Doyle, Greta Goodman, Ted Green, Michael Hall, Greg Hradsky, Thomas Kravitz, Vincent Matanoski, William McDowell, Anthony Olsen, Larry Park, Michael Plasay, Merle Pokempner, Fred Rutledge, Judith Samkoff, Keith Slifer, and Betty Ward. Dr. Margaret Nellis is a Post-Doctoral Fellow on the unit. The behavioral program and the methods adopted for its implementation were designed and operationally proven in collaboration with Cleeve S. Emurian.

References

1. Thorndyke, P.W. and M.G. Weiner, *Improving Training and Performance of Navy Teams: A Design for a Research Program*, Santa Monica, RAND (1980).

2. Hare, A.P., *Handbook of Small Group Research*, New York, Free Press of Glencoe (1976).

3. Emurian, H.H., "A multiple task performance battery presented on a CRT," JSAS *Catalog of Selected Documents in Psychology*, 8 (1978) 102.

4. Morgan, B.B. and E.A. Alluisi, "Synthetic work: Methodology for the Assessment of Human Performance," *Perceptual and Motor Skills, 35* (1972) 835–845.

5. Emurian, H.H., J.V. Brady, J.L. Meyerhoff, and E.H. Mougey, "Behavioral and Biological Interactions with Confined Microsocieties in a Programmed Environment," J. Grey, and L.A. Hamden, (Editors), *Space Manufacturing, 4*, New York, American Institute of Aeronautics and Astronautics (1981) 407–421.

6. Emurian, H.H., C.S. Emurian, G.E. Bigelow, and J.V. Brady, "The Effects of a Cooperation Contingency on Behavior in a Continuous Three-person Environment," *Journal of the Experimental Analysis of Behavior*, 25:3 (1976) 293–302.

7. Brady, J.V. and H.H. Emurian, "Behavior Analysis of Motivational and Emotional Interactions in a Programmed Environment," R. Dienstbier and R. Howe (Editors), *Nebraska Symposium on Motivation*, Lincoln, University of Nebraska Press (1979).

8. Emurian, H.H., C.S. Emurian, and J.V. Brady, "Effects of a Pairing Contingency on Behavior in a Three-person Programmed Environment," *Journal of the Experimental Analysis of Behavior*, 29 (1978) 319–329.

9. Emurian, H.H. and J.V. Brady, "Small Group Performance and the Effects of Contingency Management in a Programmed Environment: A Progress Report," JSAS *Catalog of Selected Documents in Psychology*, 9:58 (1979).

10. Emurian, H.H., C.S. Emurian, and J.V. Brady, "Appetitive and Aversive Reinforcement Schedule Effects on Behavior: A Systematic Replication," *Basic and Applied Social Psychology*, 3:1 (1982) 39–52.

11. Eberhart, J.A., E.B. Keverne, and R.E. Meller, "Social Influences on Plasma Testosterone Levels in Male Talopoin Monkeys," *Hormones and Behavior*, 14 (1980) 247–266.

12. Bernstein, I.S., R.M. Rose, T.P. Gordon, and C.L. Grady, "Agonistic Rank, Aggression, Social Context, and Testosterone in Male Pigtail Monkeys," *Aggressive Behavior*, 5 (1979) 329–339.

13. Scarmamella, T.J., and W.A. Brown, "Serum Testosterone and Aggressiveness in Hockey Players," *Psychosomatic Medicine*, 40 (1978) 262–265.

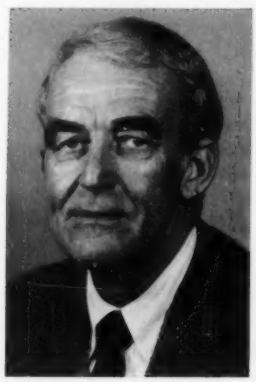
14. Elias, M., "Serum Cortisol, Testosterone, and Testosterone-Binding Globulin Responses to Competitive Fighting in Human Males," *Aggressive Behavior*, 7 (1981) 215–224.

15. Frankenhauser, M., "Psychoneuroendocrine Approaches to the Study of Emotion as Related to Stress and Coping," R. Dienstbier and R. Howe (Editors), *1978 Nebraska Symposium on Motivation*, Lincoln, University of Nebraska Press, 1979.

16. Sidman, M., *Tactics of Scientific Research*, New York, Basic Books, (1960).

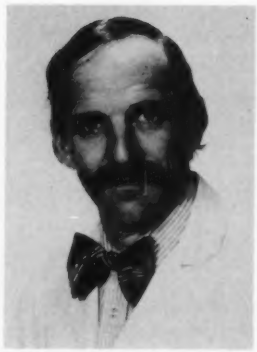
17. Turney, J.R. and S.L. Cohen, *Defining the Nature of Team Skills in Navy Team Training and Performance*, (Final Report N00014-80-C-0811, NR 170–91), Columbia, Maryland, General Physics Corporation (September, 1981).

Dr. Henry H. Emurian is a psychologist who is the Principal Investigator of this research program. He is an Assistant Professor of Behavioral Biology in the Department of Psychiatry at The Johns Hopkins University School of Medicine, and he is the Director of the Programmed Environment Research Unit.



Dr. Joseph V. Brady is an internationally recognized psychologist who is Professor of Behavioral Biology and Director of the Division of Behavioral Biology in the Department of Psychiatry at The Johns Hopkins University School of Medicine.

Dr. R. L. Ray is an Assistant Professor of Behavioral Biology. His interests include the applications of event time series techniques to the analysis of individual and team performances.



Dr. James L. Meyerhoff, M.D., is Chief of the Neurochemistry and Neuroendocrinology Branch, Department of Medical Neurosciences, Division of Neuropsychiatry, Walter Reed Army Institute of Research.

Edward H. Mougey, M.S., is Chief Endocrine Chemist, Neuroendocrine Branch, Walter Reed Army Institute of Research.

The Psychology of Decision Making

by

Daniel Kahneman
University of British Columbia
and
Amos Tversky
Stanford University

The study of decision making is shared by many disciplines, from mathematics to sociology. The normative analysis of decisions is anchored in game theory, statistics and economics; it is concerned with the rules that a decision maker must follow to optimize the expected consequences of actions taken under uncertainty and to ensure the coherence of beliefs and preferences. The normative analysis considers an idealized decision maker, who extracts all the information contained in available evidence, has a coherent set of values, and is able to find the course of action that maximizes expected utility. The psychological analysis of decision making, in contrast, is descriptive in intent: it is concerned with an analysis of people's beliefs and preferences as they are, not as they should be. Indeed, the study of the many ways in which people violate the rules of rational choice is an active focus of this research.[1,2,3]

The normative and descriptive analyses jointly provide the foundations of an emerging line of research and application, which may be called the human engineering of decisions. The aim of human engineering is to aid in the design of systems that use the abilities of human operators to good effect, but also offer as much protection as possible against the characteristic weaknesses of these operators. Principles of human engineering have been applied successfully to the design of physical devices such as cockpits, computer terminals and sewing machines. The same approach is also applicable to abstract products such as computer languages or procedures for the making of decisions. A prerequisite to a systematic attempt to improve the decision making procedures of individuals and organizations is a detailed study of the relevant characteristics of the people who will apply these procedures. We briefly discuss here two lines of research that have implications for the human engineering of decision making: studies of predictions and beliefs about uncertain events and studies of preferences in choices between risky prospects.

Predictions and Evaluations of Probability

Many decisions depend on intuitive judgments and predictions about uncertain events such as the guilt of a defendant, the quality of a team, or the resolve of an opponent. There are important similarities between such intuitive judgments and the perception of the size or distance of seen objects. Like perception, intuitive judgment is not a deliberate activity; there is little voluntary control of it. Judgment and perception are both affected by cues of imperfect validity, which yield valuable information but also produce characteristic errors. For example, the apparent distance of an object is determined in part by its clarity. This cue has some validity because in any given scene the more distant objects are seen less sharply than nearer objects, but reliance on clarity inevitably produces overestimates of distance when visibility is poor and underestimates when visibility is exceptionally good. The cues that determined judgements of probability have similar effects. Finally, there are in judgment as well as in perception illusions that remain compelling even when the observer knows the true state of affairs.

For an example of a judgmental illusion, consider an individual, who was described by a former neighbor as follows: "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure and a passion for detail." Now rank the probabilities of several occupations in which Steve could be engaged (e.g., farmer, salesman, airline pilot, librarian). There is in fact no correct way to perform this task without additional information about the procedure by which Steve was selected and the reliability of the neighbor's description.

However, intuition leaps forward unhindered by such considerations. Most readers of this description report a compelling impression that Steve is more likely to be a librarian than either a farmer or an airline pilot. This judment is dominated by the relative similarity of the personality description to the different professional stereotypes. Very few readers spontaneously bring to mind a critically important fact: that there are in the population of men many more farmers than librarians. The strong intuition that Steve "must be" a librarian qualifies as a genuine judgmental illusion because it resists such reminders of contrary arguments. The example illustrates a feature of intuitive judgment that has been amply documented by formal research: intuitive predictions tend to be dominated by the specific features of the individual case, and relevant statistical considerations are often neglected. In particular, predictions are insufficiently sensitive to the quality and reliability of the information on which they are based, and to the prior probabilities of possible outcomes.

Related biases are observed when people predict by making a "best guess" of the value of an uncertain quantity, such as the sales of a product or the time required to complete a project. The context of planning is especially rich in examples of intuitive judgments that ignore relevant statistical considerations. In the most extreme form of this phenomenon, a case is treated as unique although it belongs to a class about which the individual has considerable knowledge. Scientists and writers, for example, are notoriously prone to underestimate the time required to complete a project, even when they have much experience of past failures to live up to planned schedules. A similar bias has been documented in engineers' estimates of the completion time for repairs of power stations. This planning fallacy does not appear to reflect only wishful thinking, and it occurs even when underestimation of duration or cost is actually penalized. It is a consequence of adopting an *internal* approach to planning and prediction, which focuses on the specifics of the problem and attempts to anticipate the details of future events. In contrast, an *external* approach treats each problem of planning or forecasting as a member of a class of similar problems. For example, an attempt to forecast the duration or cost of a project should probably be based on a study of the delays and cost overruns for similar projects, quite independently of any attempt to divine the unplanned circumstances that might affect the project at hand. An external approach to such a question suggests some characteristic questions: how long do similar projects last? how accurate are the estimates of planners? does the current project differ from others in ways that may affect its duration, or the accuracy of planners' estimates? In contrast, the characterisitc internal question is: what are the specific difficulties that are likely to arise in this particular project? Although the internal approach is useful, it often yields predictions of future events that fall outside the range of previous observations. Such predictions are prone to be wrong.

Even when the pitfalls of a purely internal approach are avoided by relating the case at hand to a distribution of similar cases, intuitive predictions are often more extreme than the evidence justifies. This error of non-regressive prediction arises when people match their predictions of future outcomes to their current impressions. It appears natural to predict that a candidate who makes a very favorable impression in a personnel interview will do equally well on the job, but such a mode or prediction is unsound. The correct regressive prediction is that the individual's performance on the job will be closer to the overall average and less remarkable than the current impression. A non-regressive prediction would only be justified if outcome measures were equally likely to depart in either direction from current impressions. In fact, although there are instances in which the outcome precisely matches one's immediate impression, and some in which it is even more extreme, the more frequent result is a regression of outcomes toward less extreme values.

Studies of Preferences

Many significant phenomena of choice can be studied by asking people for their preferences between simple gambles with specified probabilities of winning or losing specified sums of money. The answers to hypothetical questions often reveal rules that also govern choices in realistic situations, where the stakes are both significant and poorly defined, and where the uncertainties cannot be quantified as neatly as a draw from a chance urn.[4,5]

The following question, for example, elicits in most people a risk-averse preference for a sure gain over a larger gain that is merely probable:

Which would you rather have:

A. A 85% chance to win $100

B. A sure gain of $80

A large majority of people prefer the sure gain of $80, although the actuarial expectation of the gamble is higher, at $85. Risk aversion in the domain of gains is a well-known effect, which helps explain why the bank rate is generally lower than the average return on speculative investments. Because they eliminate the risk and the uncertainty that people dislike, bonds and savings accounts are competitive even when their rate of return is relatively low.

The following question illustrates a phenomenon that is not as well known, although it appears to be equally robust: risk-seeking preferences in the domain of losses.

If you had to choose between the following options, what would you prefer?

C. A 85% chance to lose $100

D. A sure loss of $80

Here most people prefer the gamble over the sure loss,

although the monetary expectation of the gamble (– $85) is lower than a loss of $80.

We have studied people's preferences in a large number of problems, involving choices between two gambles or between a gamble and a sure gain or loss. The results were well described by a formal theory, which we labeled prospect theory.[4] The theory distinguishes two phases in the choice process: an initial phase in which the acts, outcomes and contingencies are "framed," and a subsequent phase of evaluation. We shall discuss the two phases in reverse order, beginning with evaluation. Consider a prospect that yields outcome x with probability p and the status quo with probability $1 - p$. The theory associates values v(.) with outcomes, and decision weights π (.) with probabilities, such that the overall value of the prospect equals $\pi(p)v(x)$.

Prospect theory assumes that a gamble is evaluated in terms of the gains and losses that are directly associated with it. As we shall see, this seemingly reasonable mode of evaluation leads to severe difficulties because there are often several ways of framing the same consequences. The theory defines gains and losses in relation to a neutral outcome or state, which is assigned a value of zero. Although subjective values differ among individuals and among attributes, we have proposed that the value function is commonly S-shaped, as illustrated in Figure 1. For example, the difference in subjective value between gains of $10 and $20 is greater than the subjective difference between $110 and $120. The same relation between value differences holds for the corresponding losses. Another property of the value function is that the response to losses is more extreme than the response to gains: the displeasure associated with losing a sum of money is greater than the pleasure associated with winning the same amount.

Consider a gamble that offers a gain of $100 with probability p. The values of the two possible outcomes associated with this prospect are v(0) and v(100), and the value of the gamble must be contained between these extremes. How does the value of the gamble vary with the probability of winning? The answer depends on the *decision weight* π (p) associated with the probability p. It is commonly agreed that the rules of rational choice require the value of the prospect to be proportional to p. For a rational decision maker, then, the decision weight π (p) is simply equal to p. It is intuitively obvious, however, that the difference between certainty and a probability of .99, or between impossibility and a low probability of .001, loom much larger than comparable differences in the intermediate range of probability. The qualitative properties of decision weights are illustrated in Figure 2. Small probabilities are overweighted, but intermediate and high probabilities are relatively underweighted. The latter effect is more pronounced than the former, so that π (p) + π (1 – p) < 1.

We commented earlier on the prevalence of risk-averse preferences in the domain of gains and of risk-seeking preferences in the domain of losses. When the choice is between a sure consequence (gain or loss) and a more extreme one that is merely probable, the uncertain outcome is relatively underweighted (see Figure 2) and the discrepancy between the outcomes is evaluated in a relatively shallow region of the value function (see Figure 1). Both effects enhance the relative attractiveness of a sure gain over a larger but risky gain, and therefore contribute to risk aversion. The same factors also increase the relative aversiveness of a sure loss, and thereby support risk-seeking preferences for gambles in the domain of losses. It should be noted, however, that the pattern of risk aversion in gains and risk seeking in losses is prevalent only in choices between sure things and gambles

Figure 1

A hypothetical value function.

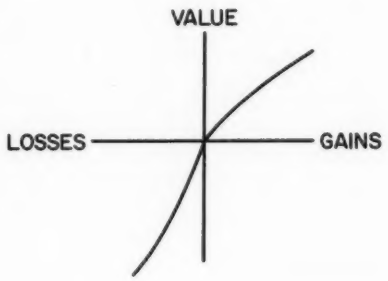
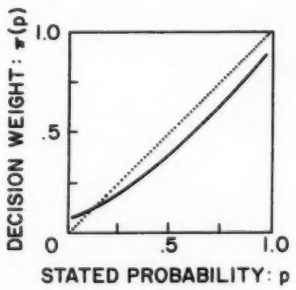


Figure 2

A hypothetical weighting function.



with substantial probabilities of winning and losing. The pattern is often reversed when the probabilities are small: the purchase of lottery tickets represents risk seeking in the domain of gains and the purchase of insurance represents risk aversion in the domain of losses. The overweighting of small probabilities (see Figure 2) increases both the attractiveness of the lottery ticket and the aversiveness of the risk for which insurance is offered.

Framing Effects

We shall now illustrate how the choices described by prospect theory violate normative rules of rational decision. The demonstrations have the following general form: we construct alternative versions of a choice problem, which most people recognize as essentially identical, but which nevertheless yield different preferences when evaluated by the rules of prospect theory. We call such reversals of preference framing effects.[4]

Our first example illustrates the effect of describing the outcomes of an action as gains or as losses.

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
>
> If Program A is adopted, 200 people will be saved.
>
> If Program B is adopted, there is ⅓ probability that 600 people will be saved and ⅔ probability that no people will be saved.
>
> Which of the two programs would you favor?

The majority choice in this problem is risk averse: the prospect of certainly saving 200 lives is more attractive than a risky prospect of equal expected value. Now consider another version of the problem, in which the same cover story is followed by new descriptions of the consequences of the two options:

> If Program A′ is adopted, 400 people will die.
>
> If Program B′ is adopted, there is ⅓ probability that nobody will die, and ⅔ probability that 600 people will die.

In this version, a large majority of respondents exhibit a risk seeking preference for option B′ over A′. It is readily seen, however, that the two problems are equivalent, and that they differ only in the location of an implicit reference point. In the first version, the consequences of the programs are formulated as gains (lives saved) relative to reference state in which 600 persons die. The second version expresses the same consequences as to losses relative to a state in which no one dies.

Many different problems of the same general form have been presented to thousands of respondents. The problems concerned money (sometimes real choices among gambles for significant stakes), hours of pain, years of post-treatment survival, and human lives. The populations of respondents varied in sophistication, from naive undergraduates to public health experts and decision analysts. The framing effect has proved to be extraordinarily robust; in all problems people are risk seeking in the version concerned with losses and risk averse when they consider the same options as gains relative to another reference point.

Two important conclusions can be drawn from these findings. The first is that people do not easily penetrate beyond the surface in interpreting choice problems and do not recognize the identity of problems that are in fact equivalent. As a result, preferences tend to be dominated by the terms in which the outcomes are initially described and can be altered by inconsequential variations in the formulation of problems. This weakness is noteworthy because in other contexts people spontaneously recode messages in an abstract form, so that, for example, we may not recollect after a few seconds whether we were told that "Mary kicked John" or that "John was kicked by Mary." The second conclusion from the demonstrations of framing effects is that *only* a comprehensive description of outcomes can provide for a coherent system of preferences because only such a description is unique, and therefore resistant to framing. When a more restricted perspective is adopted, variations of reference points will cause the same effective consequences to be evaluated in different regions of the value function (see Figure 1), and inevitably produce inconsistencies.

Framing effects can be used in a similar vein to demonstrate that the use of decision weights (see Figure 2) also leads to inconsistencies. For example, consider protective actions that reduce or eliminate the probability of a hazard at some cost. The shape of the π function in the range of low probabilities suggests that a protective action which reduces the probability of a harm from 1 percent to zero will be valued more highly than an action that reduces the probability of the same harm from 2 percent to 1 percent. However, there are many situations that can be framed in both ways. For example, an insurance policy that covers fire but not flood could be evaluated either as full protection against the specific risk of fire or as a reduction in the overall probability of property loss. There is evidence to suggest that insurance will appear more valuable in the former framing than in the latter, although such a difference cannot be justified.

The most interesting framing effects are produced by the spontaneous coding of situations, rather than by the passive acceptance of the verbal formulation of a problem. Consider the following problem:[5]

> Imagine yourself on your way to a play with a pair of tickets for which you have paid $40. On entering the theater you discover that you have lost the tickets. Would you pay $40 for another pair of tickets? Now imagine that you are on your way to the same play

without having bought tickets. On entering the theater you discover that you have lost $40 in cash. Would you now buy tickets to the play?

The two situations are objectively identical: in both cases you are poorer by $40 than you were earlier, and you face the same decision of whether or not to pay $40 to see the play. However, most people report that they would be more likely to buy tickets if they lost the cash than if they had lost tickets, perhaps because they feel in the latter case that they spend $80 to buy tickets that are worth only $40. It is difficult to find rational justification for such framing effects.

Observations of the type that we have sketched here define specific challenges for the human engineering of decisions. Because many important decisions depend on intuitive judgments of probability, it is necessary to devise training procedures and decision aids that will counteract predictable biases of such judgments. It is also necessary to protect decision makers against framing effects, and against the distortions induced by the peculiarities of the value function and of decision weights.

References

1. Einhorn, H. J., and R. M. Hogarth, "Behavioral Decision Theory: Processes of Judgment and Choice," *Annual Review of Psychology*, 32 (1981) 53–88.

2. Slovic, P., B. Fischoff, and S. Lichtenstein, "Behavioral Decision Theory," *Annual Review of Psychology*, 28 (1977) 1–39.

3. Kahneman, D., P. Slovic, and A. Tversky, (Editors) *Judgment under Uncertainty: Heuristics and Biases*, New York, Cambridge University Press (1982).

4. Tversky, A. and D. Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, 211 (1981) 453–458.

5. Kahneman, D. and A. Tversky, "The Psychology of Preferences," *Scientific American*, 246 (1982) 166–173.

Authors

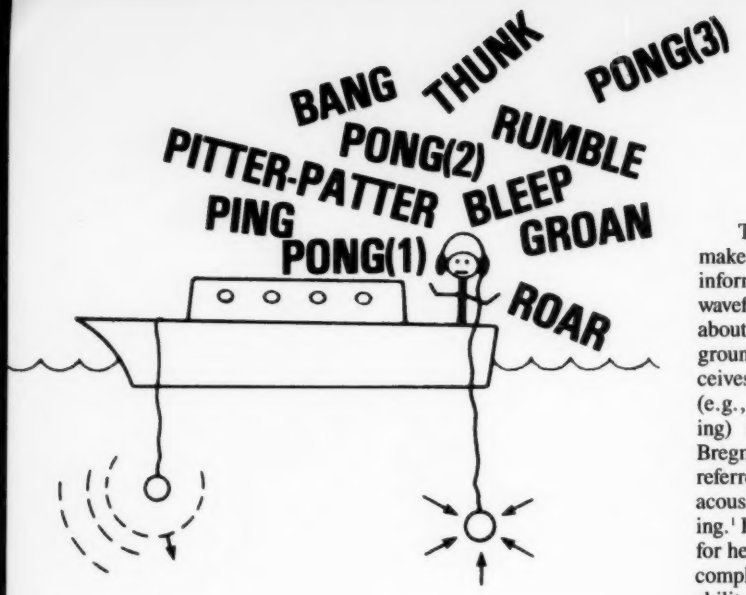
Dr. Amos Tversky teaches psychology at Stanford University, and Dr. Daniel Kahneman teaches psychology at the University of British Columbia in Vancouver, Canada. Their collaboration in the study of judgment and choice started in 1969, when both were associated with the Hebrew University in Jerusalem. They were jointly awarded in 1982 the Distinguished Scientific Contribution Award of the American Psychological Association. Their research has been supported by ONR since 1978.

Profiles in Science

Dr. Amos Tversky is Professor of Psychology at Stanford University and has been a Principal Investigator for the Office of Naval Research (ONR) since 1978 when he joined the faculty at Stanford. Professor Tversky and his colleague, Professor Daniel Kahneman, are renowned for their joint research which introduced new and precise methods for studying judgment and decision making. Their understanding of the role of heuristics in judgments of uncertain events has added new dimensions to the field of cognitive processes. Both men were honored for distinguished scientific contributions in 1982 by the American Psychological Association. The definitive text in the field of heuristics and inference in decision making, *Judgment under Uncertainty: Heuristics and Biases,* was compiled and edited by Tversky, Kahneman, and Professor Paul Slovic in 1982 with the support of ONR. The book presents the research findings not only of Tversky and Kahneman but of others in the field, including several other ONR investigators.

Beginning in 1961 with his doctoral dissertation at the University of Michigan and continuing to the present, Tversky has made noteworthy contributions to the study of attention and mathematical psychology. He has studied the perception of similarity, its impact on choice, and the representation of similarity data. Realizing both the strengths and limitations of multidimensional scaling models, he developed the contrast model, which expresses the perceived similarity of objects as the result of a feature-matching process rather than as a metric distance between points.

Professor Tversky is on the editorial board of five journals, including the *Journal of Mathematical Psychology* and the *Journal of Economic Behavior and Organization,* and he is a Fellow of the American Academy of Arts and Sciences.



Human Classification of Complex Sounds

by

James H. Howard, Jr.
The Catholic University of America
and
John J. O'Hare
The Office of Naval Research

The human listener possesses a remarkable ability to make sense out of multiple and diverse arrays of acoustic information. At any particular instant the complex acoustic waveform arriving at the ear will likely contain information about a variety of individual sound sources as well as background noise. Despite this, the experienced listener perceives sounds from each of the separate sources individually (e.g., the dog barking, a Bach cantata, the telephone ringing) rather than as a nonsensical hodgepodge. Albert Bregman, a perceptual scientist at McGill University, has referred to this ability—the act of sorting our perceived acoustic world into separate sources—as auditory streaming.[1] He has argued convincingly that people are specialized for hearing acoustic *sources*, not acoustic *components*. The complexity of the perceptual processes which underlie this ability has been underscored by the difficulties encountered in the Defense Advanced Research Projects Agency project to develop automatic classifiers for speech.[2]

This impressive human pattern recognition ability has been exploited in a number of Navy applications which involve acoustic signal processing. In both *passive sonar* and *ship silencing* operations, complex acoustic data are analyzed for their tactical significance. In passive sonar, sounds from the underwater environment are recorded on hydrophones, processed, and presented on either visual or acoustic displays. Both long-duration steady state and brief-duration transient signals occur from a variety of sources as illustrated in Figure 1. The task of the sonar operator in this context is to distinguish the sources of radiated noise from each other as well as from the background of ambient and platform-generated noise. Ultimately, the identity (friend or foe?) and intention (threat?) of these noise sources must be determined. The ability of experienced sonar technicians to make these decisions is legendary. In ship silencing, the objective is to reduce the radiated noise produced by a ship not only in magnitude but also in detectability/classifiability. Unwanted sources of noise must be identified and eliminated to reduce the vulnerability of the vessel to early detection by enemy sonar. The sonar and the ship silencing problems represent opposite sides of the same perceptual coin. On the passive sonar side, one searches for cues to facilitate detection and classification whereas on the ship silencing side such cues are eliminated to make detection and classification difficult.

Despite the significance of the auditory recognition problem, surprisingly little research has investigated the ability of human listeners to identify complex nonspeech sounds. In the present paper, we summarize some of our recent research which has investigated the perceptual processes that underlie auditory pattern recognition.

Contributions of Human Auditory Research

We live in an age of technology—an age of ever-increasing automation. A natural question to pose in this context is "why study basic human perceptual capabilities when the human is likely to be automated out of the system?" After all, intelligent, knowledge-based computers are reliable, adaptable systems which are especially well-suited to tedious surveillance and monitoring tasks. In our view, the answer to this pragmatic question lies in recognizing the value of basic human research to the automation development effort. Few current pattern recognition systems are fully automated. Rather, most systems incorporate some combination of human operator and intelligent computer in an interactive environment. An optimized system will achieve a division of labor which is best suited to the capabilities of both the human and the computer.

Basic perceptual research can contribute to this in two ways. First, once the processes which underlie the human's listening capability are understood, preprocessing performance aids can be developed to facilitate the operator's task. Second, in many cases the very design of intelligent, automated systems depends in some way on understanding how a human performs the task. For example, H. Penny Nii and Edward Feigenbaum of Stanford University have developed an intelligent, knowledge-based expert system called HASP (SU/X in its earlier versions), for use in machine-aided interpretation of data from an extensive underwater surveillance system.[4,5] The system is based on the HEARSAY-II speech understanding system[6] and is designed to identify and track multiple ship targets by monitoring spectral data over time—much as a sonar analyst monitors a sonogram display. Nii and Feigenbaum have argued convincingly that traditional statistical signal processing methods are ineffective in this complex environment; what is critical is that the system has a knowledge base or level-of-expertise similar to that of the experienced human technician. Given this expertise, the HASP system can use some of the rules or heuristics employed by experienced analysts in data interpretation.

Figure 1

Schematic representation of various sounds in the ocean (reference 3).

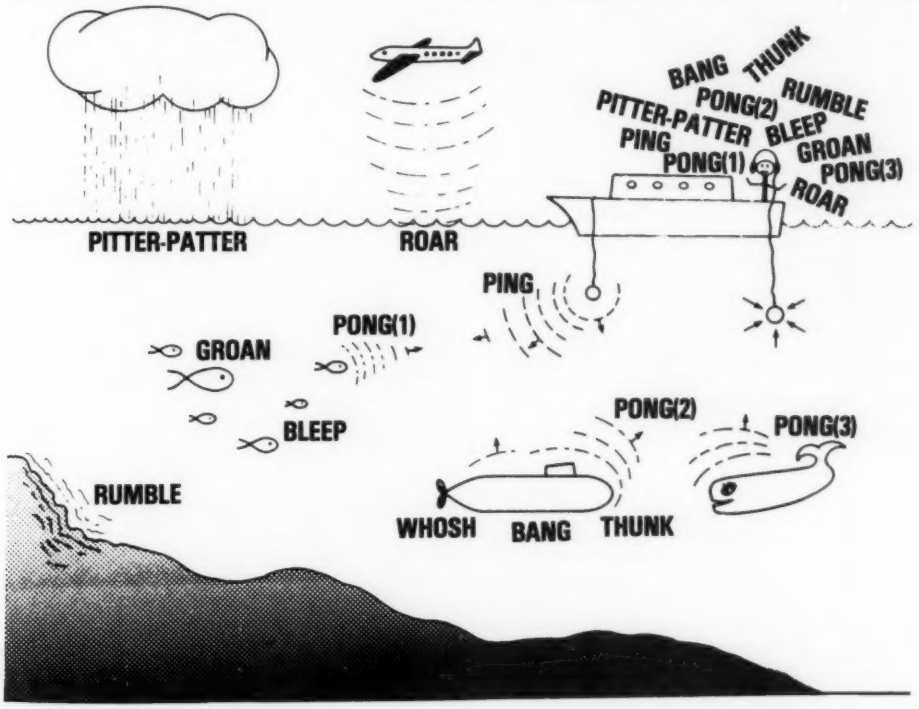
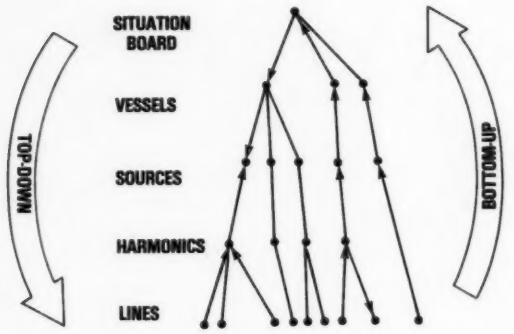


Figure 2

Hierarchy of interpretation levels for underwater sounds.



As in other expert systems, the knowledge base for this system was developed through an extensive analysis of the procedures followed by expert human technicians. HASP has achieved impressive success with large amounts of real data at poor signal-to-noise ratios. In its current implementation, the raw data input to the system—equivalent to sonogram lines and line parameters—are encoded by human operators for input to HASP. Basic research such as that described below will be of use in developing fully- or partially- automated systems for sonar analysis.

Auditory Perceptual Processes

Our approach is based on the assumption that both bottom-upward or data-driven and top-down or knowledge-driven processes are important in the perception of complex nonspeech sounds. *Bottom-up* processing is based exclusively on the information available in the acoustic waveform. Specific cues or features are present which suggest a particular category or source event. In contrast, *top-down* processing occurs when the perceiver uses prior knowledge to generate expectancies of what is likely to occur. This information is not available in the signal itself, but is applied by the perceiver in interpretation. Figure 2 illustrates bottom-up and top-down processing as it occurs in the HASP system. Several levels of interpretation are proposed which range from low-level spectral line parameters to high-level multi-ship configurations. Bottom-up processing proceeds up through the interpretation hierarchy (i.e., certain line parameters imply particular harmonics which imply source events, etc.), whereas top-down processing proceeds from higher to lower levels (i.e., the expectation that a particular ship was present suggests that particular sources and harmonics exist).

The most convincing evidence that human auditory perception involves both bottom-up and top-down processing is found in the speech perception literature. Logically, the "raw data" or specific sounds in continuous speech are not sufficient to account for language understanding since the "raw" input is neither complete nor unambiguous. Rather, the listener must rely on prior knowledge of the proper ordering (syntax) and the meaning (semantics) of the words in the utterance. Those elements that can be interpreted unambiguously provide constraints to facilitate the interpretation of other elements. It is common for people to "hear" missing words or to "correct" mispronounced words when they occur in continuous speech. Many recent studies have demonstrated that listeners use a number of knowledge sources in speech perception ranging from fairly specific item-to-item syntactic constraints to global semantic considerations such as the theme or title of a story. Two speech scientists, Ronald Cole and Jola Jakimik, have recently concluded that "it is not only what we hear that tells us what we know; what we know tells us what we hear."[7]

The results of a number of our recent experiments have led us to conclude that much the same thing occurs in the perception of complex nonspeech sounds.

We have examined sounds which share properties with two broad classes of natural and ship-produced underwater signals. First, we studied the perception of steady-state signals which reflect the radiated background noise of normal ship operations. The broadband hydrodynamic sounds produced by cavitating propellers are representative of this category. Second, we investigated human processing of a large class of nonsteady-state sounds produced by relatively brief-duration mechanical events such as hatch closings or pump operation. Events of this latter type generate acoustic transients. Current automatic signal-processing devices have limited utility for classification of transients, and brief-duration signals are not easily depicted on visual displays.

Our research has shown that bottom-up processes are particularly important in the perception of steady-state sounds whereas top-down processes are especially significant for the perception of most acoustic transients. The long duration of steady-state sounds provides ample cue redundancy and sufficient time for feature analyses. In contrast, transient signals are often too brief to permit an extensive analysis of their perceptual features. Furthermore, acoustic transients frequently occur in temporal succession forming more complex transient patterns. Since these patterns mirror the physical events which produced them, the order of transient components within a pattern is not arbitrary, but reflects the temporal structure of the generating events. Although the temporal structure in patterns of this sort is clearly less rigid and well specified than the syntax of language, a consistent temporal structure exists. Our research has indicated that listeners will make use of prior knowledge of this structure when classifying patterns of complex environmental transients.

To summarize, while bottom-up processing is obviously important in the perception of both steady-state and transient signals, considerable top-down processing can and does occur when listeners perceive transient patterns. Some of the evidence which supports these conclusions is summarized in the following sections.

Bottom-up Processing

Many bottom-up theories of auditory pattern recognition assume that early in processing a signal is broken down into a set of elementary properties or perceptual features. In bottom-up processing, these features form the basic building blocks of auditory recognition. A major problem, therefore, is to identify the features used by listeners in the bottom-up processing of a particular sound or set of sounds. However, since the feature extraction processes are not directly observable, the features can only be inferred from performance data.

We have employed multidimensional scaling techniques to identify the features used by listeners in processing sets of complex sounds. In one experiment, we examined sixteen simulated engine sounds which varied along four physical characteristics. Listeners were presented with all possible pairs of these signals and were asked to evaluate the degree of similarity of each pair. The resulting matrix of similarity judgments was then analyzed using multidimensional scaling. The outcome of this analysis is a decomposition of the perceptual data into a set of orthogonal dimensions with each signal represented as a point in the dimensional space. The dimensions spanning this space are interpretable as the perceptual features used by the listeners in bottom-up processing. Our scaling results suggested that listeners used three perceptual features to compare the engine sounds.[8] Furthermore, a systematic correspondence was observed between these perceptual features and some physical properties of the signals which would enable the potential confusability or distinctiveness of particular sounds to be predicted. In addition, entire sets of complex sounds can be depicted in the perceptual space to achieve economy of representation for classification training. This is illustrated in the three-dimensional plot of sixteen diesel engine sounds in Figure 3.

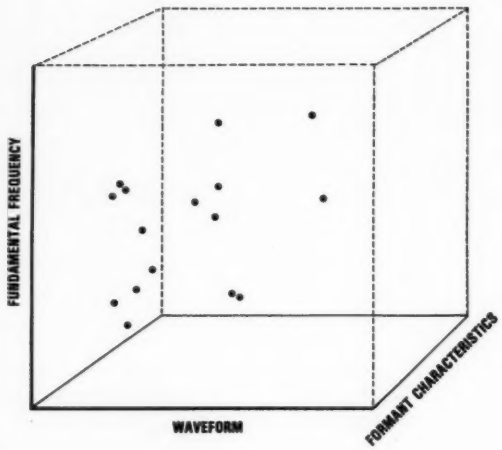
In a second experiment, we applied this method to a set of common background and radiated passive sonar signals.[9] A multidimensional scaling analysis of the data revealed a two-dimensional perceptual feature space for the sounds. Furthermore, when these perceptual features were compared to a physical analysis of the signals, it was found that listeners were using the overall shape of the low-resolution spectra, as well as the presence or absence of a low-frequency periodicity as cues. Hence, both frequency and time domain cues were used. These psychological dimensions have been verified by Mackie and his colleagues with a group of experienced watch-standers as listeners.[10]

Feature extraction involves information reduction; some information in a sound is retained, whereas other information is discarded. Ideally, the set of features should uniquely specify each sound, enhancing perceptually important differences while reducing or eliminating perceptually unimportant differences. Another issue we have addressed concerns the selection of features for signal comparison. How are the elementary units of comparison—the features—determined?

Our subsequent research has indicated that the human internalized a set of rules and criteria for feature selection rather than established a set of finely-tuned feature detectors.[11,12] These rules enabled the listener to determine what the comparison features should be in any particular listening context. In other words, for complex sounds features are not absolute, but depend on the signal context. We specified a quantitative model to predict the acoustic features which listeners use in the bottom-up processing of complex sounds. In the model, an initial low-resolution spectral analysis is carried out on the time-domain sounds. A second transformation then extracts the set of comparison features

Figure 3

Three-dimensional psychological space for 16 diesel engine sounds.



which account for as much of the variability among the spectral representations of the comparison signals as possible. This model accounted for the features employed by listeners in comparing a number of steady-state complex sounds.[12] The perceptual features extracted in bottom-up processing can be estimated with this model for steady-state sounds in many listening contexts. Once the important features are known for a set of sounds the specific confusions likely to be made in classification can be predicted. We have demonstrated the effectiveness of the model in a study involving the classification of simulated propeller cavitation sounds.[13] Once the significant psychological dimensions and their physical correlates have been identified, the results can be employed to write algorithms for pre-processors, to improve the selection and training of listeners, to identify those aspects of platform-generated noise that are readily interpreted by human listeners, and to develop and modify algorithms for automatic target classification systems.

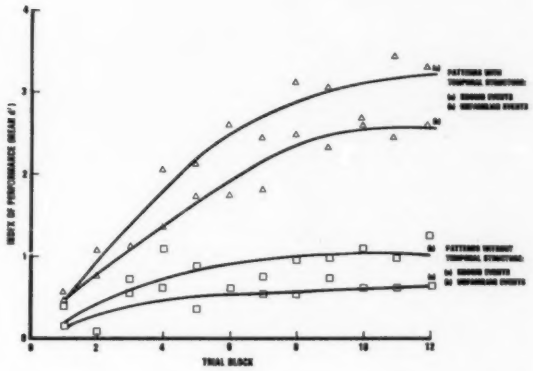
Top-down Processing

Top-down or knowledge-driven processes play a critical role in the perception of acoustic transients. We have conducted a number of experiments to investigate the role of both knowledge of temporal structure (i.e., syntactic knowledge) and knowledge of the source events (i.e., semantic knowledge) in two-alternative (target/nontarget) classification of transient patterns.[14,15] Our findings have shown that both factors can play an important role in the classification of such patterns.

The importance of temporal structure was demonstrated in an experiment where listeners classified tonal patterns. Listeners who categorized a temporally structured target set performed better than those who heard a temporally unstructured set. This difference in knowledge about the source events was demonstrated to be important in an experiment involving patterns of familiar, brief-duration complex sounds such as pipe clangs, steam hiss, and other steam- and water-related events. Semantic knowledge about those events was provided to some listeners in the form of explicit descriptions of the pattern components. The classification performance obtained for each of the four conditions in this experiment is shown in Figure 4. Results revealed that knowledge of the source events led to improved classification performance for interpretable, temporally-structured patterns. However, it was surprising to find that the same semantic knowledge impaired performance when dealing with temporally-unstructured patterns. It can be conjectured that in classifying ambiguous or unstructured sounds, dependence on prior knowledge of natural pattern structure leads to a greater number of false classifications. Knowledge of temporal structure and knowledge of the source

Figure 4

Classification performance with "real world" sounds for structured or unstructured patterns and with or without thematic information about the sounds.



events interact in an important way to influence classification performance. Individuals rely on their knowledge of the sounds that are likely to occur as well as on the specific perceptual features present in the acoustic signal.

In our recent work, we have investigated two practical implications of this finding. First, top-down processing can lead to improved performance when it helps the individual focus attention on a particular pattern component. Such *attentional focusing* can lead to an enhanced ability to resolve individual elements from multi-component acoustic patterns. Second, top-down processing may impair performance whenever it leads the listener to "hear" pattern elements which are not physically present in the acoustic signature. Although such *auditory induction* can be helpful in interpreting highly-redundant speech signals, misinterpretation can occur when listeners must detect individual components in noisy underwater sound patterns.

An important issue that we have addressed in attentional focusing concerns the ability of listeners to use the earlier pattern components as cues to listen selectively for later elements. We have carried out a series of experiments in which listeners were presented with 12-element tonal patterns at a low signal-to-noise ratio. They were required to determine which of two consecutive presentations of the pattern had the eleventh element missing and which was complete. Our findings indicated that as listeners acquired experience with the patterns, they were able to use the earlier components as cues to predict the later elements.[16] This resulted in an increase in sensitivity to the expected component with a contrasting inability to hear unexpected elements. Furthermore, this sensitivity changed moment to moment as a function of the attentional cues provided by earlier pattern components.

In another series of experiments we have examined how auditory induction can lead to impaired performance when listeners are required to detect signals presented under various listening conditions. Listeners were asked to detect signals embedded in noise bursts which were preceded and followed by an acoustic context designed to be either consistent or inconsistent with the to-be-detected signal.[17] Our results showed that the detection of "in-context" signals was associated with an increased false alarm rate and lowered sensitivity compared to "out-of-context" signals. In other words, sufficient top-down processing occurred for listeners to "hear" the in-context signals even when they were not actually present.

Conclusions

Historically, most psychoacoustic research has focused on the ability of listeners to detect isolated pure tones of relatively brief-duration. Although this research has led to significant advances in our understanding of energy detection mechanisms in the peripheral auditory system, it has contributed little to our understanding of how human listeners classify complex environmental sounds. The primary difficulty with this approach is that it necessarily focuses on bottom-up and disregards top-down processing. In the contextually rich acoustic environment of underwater sounds, listeners employ a variety of knowledge sources to identify a sound pattern. The experienced technician knows a great deal about the sound sources likely to occur, and this knowledge plays a major role in determining what is heard. A fuller understanding of these bottom-up and top-down processes is essential for the ultimate development of partially- and fully-automated systems for passive sonar.

References

1. Bregman, A.S., "The Formation of Auditory Streams," J. Requin (Editor), *Attention and Performance VII*, Hillsdale, N.J., Erlbaum (1978).

2. Lea, W.A. and J.E. Schoup, "Specific Contributions of the ARPA SUR Project," W.A. Lea (Editor), *Trends in Speech Recognition*, Englewood Cliffs, N.J., Prentice Hall, 1980.

3. Bisson, A.E., "Pattern Recognition in Ocean Acoustics," D.J. Getty & J.H. Howard, Jr. (Editors), *Auditory and Visual Pattern Recognition*, Hillsdale, N.J., Erlbaum (1981).

4. Nii, H.P. and E.A. Feigenbaum, "Rule-based Understanding of Signals," D.A. Waterman & F. Hayes-Roth (Editors), *Pattern-directed Inference Systems*, New York, Academic Press, 1978.

5. Nii, H.R., and E.A. Feigenbaum, "Signal-to-symbol Transformation: HASP/SIAP Case Study," *The AI Magazine* (Spring 1982) 23–36.

6. Erman, L.D., F. Hayes-Roth, V.R. Lesser, and D.R. Reddy, "The HEARSAY-II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty," *ACM Computing Surveys*, 12 (1980) 213–253.

7. Cole, R.A. and J.A. Jakimik, "Understanding Speech: How Words Are Heard," G. Underwood (Editor), *Strategies of Information Processing*, New York, Academic Press (1978).

8. Howard, J.H., Jr. and E.B. Silverman, "A Multidimensional Scaling Analysis of 16 Complex Sounds," *Perception & Psychophysics*, 19 (1976) 193–200.

9. Howard, J.H., Jr., "Psychophysical Structure of Eight Complex Underwater Sounds," *Journal of the Acoustical Society of America*, 62 (1977) 149–156.

10. Mackie, R.R., C.D. Wylie, R.R. Ridihalgh, T.E. Shultz, and M.L. Seltzer, *Some Dimensions of Auditory Sonar Signal Perception and Their Relationships to Target Classification* (Technical Report 2723–1). Goleta, California: Human Factors Research, Inc., 1981 (AD A102598).

11. Howard, J.H., Jr. and J.A. Ballas, "Feature Selection in Auditory Perception," D.J. Getty & J.H. Howard, Jr. (Editors), *Auditory and Visual Pattern Recognition*, Hillsdale, N.J., Erlbaum (1981).

12. Howard, J.H., Jr., "Feature Selection in Human Auditory Perception," *Pattern Recognition*, 15 (1982) 397–403.

13. Howard, J.H., Jr., and J.A. Ballas, "Perception of Simulated Propeller Cavitation," *Human Factors*, (in press).

14. Howard, J.H., Jr. and J.A. Ballas, "Syntactic and Semantic Factors in the Classification of Nonspeech Transient Patterns," *Perception & Psychophysics*, 28 (1980) 431–439.

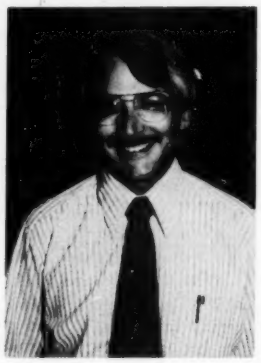
15. Howard, J.H. Jr., and J.A. Ballas, "Acquisition of Acoustic Pattern Categories by Exemplar Observation," *Organizational Behavior and Human Performance* 30 (1982) 157–173.

16. Howard, J.H., Jr., A.J. O'Toole, R. Parasuraman, and K.B. Bennett, "Pattern-directed Attention in Uncertain Frequency Detection," *Perception & Psychophysics, (in press).*

17. Bennett, K.B., R. Parasuraman, J.H. Howard, Jr., and A.J. O'Toole, *Auditory Induction of Discrete Tones in a Signal Detection Task* (Technical Report ONR-83-23). Washington, D.C., The Catholic University of America, Human Performance Laboratory, (August, 1983).

Authors

Dr. John J. O'Hare is a Scientific Officer in the Life Sciences Division of ONR. His areas of primary interest are visual and auditory perception, human-computer interaction, and human factors in software engineering.



Dr. James H. Howard, Jr. is Professor of Psychology and Director of the Human Performance Laboratory at the Catholic University of America, Washington, D.C. His current research interests are in auditory and visual pattern recognition and interactive computer image processing. He has been a recipient of the Award for Scientific Achievement in the Behavioral Sciences (Washington Academy of Sciences, 1977).

Cognitive Psychology

by

Marshall J. Farr
Office of Naval Research

In recent years, the Personnel and Training Group has been putting increasing emphasis on studies based on the perspective of cognitive psychology. For years, this perspective was poorly represented in human experimental psychology, which was dominated by the Behaviorist doctrine that only observable responses (such as error rates and error types in the learning of a list of nonsense syllables) were "legitimate" for psychology to study. Under the Behaviorist perspective, the human information-processing system, since it was not directly observable, was treated as an inaccessible "black box" and essentially ignored. However, that same black box is now the prime target of the cognitive psychologist, who is primarily concerned with how people take in, transform, manipulate, store and retrieve information. Theorizing about the human information-processing system amounts to trying to open and explore the black box. One conceptual flow depiction of the system is shown in Figure 1, which assumes a series of processing mechanisms that accept information about the environment, perform general central-processing operations, and control motor output. Readers familiar with computers and artificial intelligence should see some resemblance to how a computer program might be organized. Certainly modern cognitive psychology has been influenced by developments in computer science.

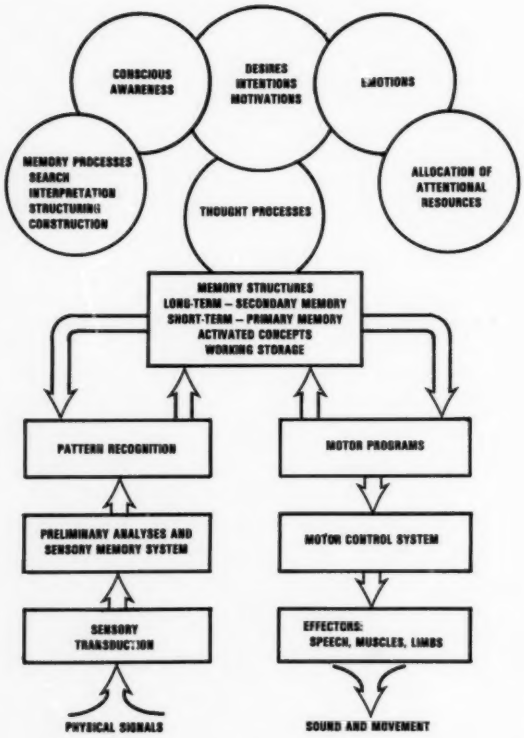
ONR's Program

ONR's cognitive psychology program constitutes an effort to identify and understand the microprocesses—the indivisible component processes—underlying the different stages of human information handling. Dr. Robert Sternberg, an ONR contractor at Yale University, has been performing what is probably the most ambitious and pioneering work in this area as he tries to formulate an information-processing theory of human reasoning and intellectual behavior. He has applied his "componential analysis" approach to studying analogical-reasoning problems of the type found in most IQ tests, and he has described his work in the pages of this magazine in some detail (see *Naval Research Reviews*, Vol. XXXIII, No. 1, pp. 48-68). The search for the basic components of human cognitive or intellectual behavior is different from the traditional psychometric approach of constructing intelligence-test items to tap ability factors which derive solely from the statistical technique of factor analysis. In terms of factor analysis, broad concepts like intelligence, and so-called "primary-ability" factors like logical reasoning, visual memory, perceptual speed and verbal fluency (to name just a few) are statistical abstractions or constructs with no basis in theory. The component approach used by Sternberg (and a number of other cognitive researchers being supported by ONR) is aimed at more than simply isolating and precisely measuring basic human information-processing operations. We are concerned with how these processes contribute to individual differences in test-like tasks, in learning, and in real-world performance in the wide variety of tasks found in the Navy. We're interested in how these microprocesses combine with each other in the service of different tasks. And we are concerned with whether or not these processes are trainable. We are pursuing the information-processing approach to intelligent behavior and complex performance along a number of related fronts.

Work such as Sternberg's has shown that one of the ways in which people can differ is with regard to strategies employed to solve a problem or learn a task. Strategies are reflected in the choice of components to use and in the particular ways the components chosen are combined by the person. Sternberg uses the term "meta-components" to mean the higher-level, executive-decision processes that determine what strategies are called into play. A number of

Figure 1

A modern version of the conventional flow chart of the human information processing system. The basic components are a series of processing mechanisms that take in information about the environment, perform general central processing operations, and control motor output.



interesting findings have emerged from studying people's differing strategies:

1. As they gain experience with a given type of problem, they may change strategy either consciously or unconsciously.
2. Among adults, the major source of individual differences in performance on analogy problems can be found in the time taken to execute various components, and the accuracy achieved at each step.
3. Metacomponents seem to be more importantly involved in general intelligence than are components. IQ tests may be deficient because they emphasize various aspects of problem *solution* (such as accuracy or correctness) rather than highlighting the persons' decisions about *how* the problem should be solved. Consider the role of stimulus encoding in intelligence in general and analogical reasoning in particular. Contrary to what one would intuitively expect, people with higher IQs tend to spend *more* time encoding than people with lower IQs. A reasonable explanation seems to be found at the metacomponent level: better problem solvers spend more time on initially encoding the terms of the problem because successful encoding then allows them to carry out, more efficiently, the operation they must later perform upon these encodings. An example here is that of word problems that need to be encoded or transformed into algebraic notation before they can be solved. Once a problem is correctly encoded, the solution is often obvious, or at least much simpler.

Dr. Richard Snow of Stanford University, a recent ONR investigator using the component approach to information processing, has also tried to find strategies which differentiate low-aptitude from high-aptitude individuals. By analyzing eye movements of people answering various test questions, he has found that high-aptitude processing is characterized by a two-stage procedure: it first constructs the correct answer, and then searches the response choices for the answer. The less-successful, low-aptitude processors try to do parallel processing of both item stems and alternative responses. This is induced from the fact that their eyes roam back and forth in a not-very-systematic fashion.

Once we start thinking in terms of information-processing components, strategies and metacomponents, we begin to see that individual differences in cognitive performance may be attributable to differences in the component processes, or in the strategies employing the components, or in the metacomponents. One of the chief reasons we are interested in identifying individual-processing differences lies in the practical issue of their *trainability.* It would seem that strategies and metacomponents would be more profitable targets for training than one's microprocesses. We have already begun to examine this area of information-processing

trainability. Sternberg, for example, has succeeded in improving people's abilities to solve linear syllogisms by strategy training. His continuing concern is with matching the kind of training to the component-ability patterns of the person, and discovering what it is about certain strategies that makes them more efficient than others.

Studying Real-World Tasks

Strategy training and metacomponent-level training seem to hold great promise for improving performance in specific, rather simple, well-defined kinds of tasks. But the crucial question is: what about complex real-world tasks, such as troubleshooting a sophisticated electronic system to locate and repair a malfunction, or interpreting a radarscope filled with rapidly changing patterns? Complex reasoning and problem solving typically require information not contained solely in the problem statement. Rather, real-life problems usually require drawing upon large stores of information in long-term memory. Thus, to the *processing* operations discussed above, we must add the *structural* or organizational characteristics of long-term memory. Cognitive psychologists devote a great deal of attention to discovering how we represent knowledge in memory, and how we access our representations for purposes of solving problems and performing tasks.

One major technique for addressing these issues is that of "protocol analysis," which is based on having experimental subjects, during the actual process of doing a task or problem, describe what they are doing and why in detail. The subjects think aloud, and psychologists try to infer their strategies and the nature of their knowledge representations from their accounts. The other major approach for studying complex cognitive activity is that of simulation modeling. Here, we write computer programs that mimic cognitive operations. Beginning with the computer modeling of chess playing, cognitive scientists have now formulated simulation models for concept formation, and for language and text comprehension. The essence of the advantage of this computer-simulation technique is that in order to program simulated human behavior for a computer one has to know something quite explicitly about the behavior being modeled. All steps in a computer program have to be clearly and completely specified if the program is to supply an output that accurately matches the behavior of human subjects. The advantage here is that the computer serves to check the adequacy of the theoretical assumptions that the psychologist believes will account for the psychological process under study. Note also that an adequate simulation of a process can be the basis for designing an intelligent, computer-based training system.

Our cognitive psychology program is also aimed at probing into the subtle operations that underlie skilled performance in real-world tasks demanding complex information processing. A major goal of this research is to build models of skilled problem solving and expert performance which directly represent the relevant cognitive processes and structures. These models will incorporate parameters reflecting the characteristics of the individual problem solver or expert, his or her training history, and the problem-solving environment.

ONR has been pursuing a research program in the area of complex cognitive behavior for several years now. One of our initial areas of interest was the collection, comprehension, and analysis of information, primarily in the form of text—the kind of task done by a military analyst, for example. The research has spanned subjects as varied as topic identification, making inferences from diverse information sources, and the relationship between verbal and graphic material. The results so far suggest that recently developed theories of semantic structure have strong potential for characterizing a variety of analysis and comprehension tasks. To explore this potential, we have started a small research thrust on how best to convey procedural knowledge via text and other media.

Another important emphasis in this ongoing program is on human planning and problem solving. Here we have focused on, or are focusing on, areas such as software design, physics, radiology, the use of maps in route planning, arithmetic, geometry, and the interpretation of engineering drawings. As the evidence accumulates from our research across these areas, it strongly supports the view that what gives the expert a unique advantage over the novice is his or her memory-storage structure—how richly the facts and procedures and relationships relevant to a domain are represented in the head. The expert has condensed or chunked many large blocks of related information in a meaningful way, so that he or she can quickly and successfully search through a well-organized, well-encoded memory. Our interest in what it takes to make and to be an expert is motivated by our desire to determine whether or not we can, through special training approaches and techniques, find short-cuts to expertise.

Conclusion

The importance of this cognitive research to the military should be clear. The kinds of tasks we are now considering—such as map interpretation, equipment troubleshooting, military-intelligence analysis, and computer programming—are confronted daily, and in many cases are becoming more demanding. Expertise in these sorts of tasks is often restricted to small numbers of individuals who appear to be exceptional. That expertise is usually acquired with considerable difficulty and in many cases years of practice and experience. Furthermore, the cognitive processes involved in these tasks are commonly not available to the expert's own awareness. Therefore, even though we can infer some of the expert's processes and structures through

protocol analysis, we do not yet have the necessary explicit knowledge of his representational patterns and strategic operations that will allow us to improve the instruction of novices or design clever job aids. Demands for high levels of expertise are increasing in the military; qualified instructors are in short supply in many occupations as are personnel in high-technology ratings. The current research emphasis on complex information-processing is aimed at providing explicit, formal theories of the crucial cognitive components of required skills and knowledge. These theories will constitute the foundations for selection of individuals with high potential, for better diagnosis of cognitive strengths and weaknesses, and for training methods which can both shorten the time required for skill acquisition and mastery, and produce a more competent individual.

The following article by Dr. Earl Hunt provides a concrete example of the ways in which the cognitive psychologist approaches the black box of human information processing and related topics, such as the assessment of "intelligence."

Author

Dr. Marshall J. Farr is head of the Personnel and Training Research Group at ONR. His research has been primarily in the fields of cognitive and instructional psychology, human information processing, and psychometrics. He is currently editing a book with Dr. Richard Snow on the relationship between cognition and affect.

Intelligence and Mental Competence

by

Earl Hunt
The University of Washington

Formal methods for evaluating "intelligence" have been a part of military personnel classification since World War I. Intelligence testing itself has become a part, and often a controversial part, of our culture. Although intelligence tests are part of the technology of personnel management, the scientific concept of intelligence has remained murky. Beginning about 1970, psychologists interested in cognition began to take a new look at the nature of individual differences in mental competence, a topic that subsumes intelligence testing but certainly is not identical to it. This article presents some of the conclusions that have been reached. In the interests of brevity, selected experimental studies will be presented as illustrations of the major points. For more extensive reviews, see references 1 and 2.

It is easier to say what intelligence is not than to say what it is. Intelligence is not a unitary attribute of a person as height is. A better analogy is to athletic ability. While it is intuitively reasonable to speak of a "good athlete," athletic ability is really a collection of skills in manipulating the body. A good athlete must possess certain minimum anatomical and physiological properties, a general style for approaching physical activities, and finally, a good athlete must know the sport that is being played. Thinking places similar requirements on people. Thinkers manipulate mental representations of external problems, using different classes of mental skills.

What are these classes? First the internal representation must be constructed. How is the problem to be represented to oneself? What are the essential features, and what are the inessential ones? Some technique for problem solving must then be chosen. Ideally, all this involves is recall; the easiest problems to solve are those whose answers are known. In other cases (e.g. doing mental arithmetic) the answer may not be known, but the thinker may know procedures that are guaranteed to get the answer to problems "like" the one at hand. Such knowledge can only be used if the thinker has some way of classifying the present problem as suitable for attack by this or that problem solving method. Methods themselves vary in generality. Some methods will work on many problems but are unwieldy in almost every application. Other methods are very efficient but applicable only in special situations. Should a good thinker use general or special methods? Finally, once a thinker has decided how to attack a problem, executing the chosen method requires the execution of a number of information processing steps, analagous to limb movements or perceptual identifications in athletics. The solution method must not demand more from a person's information processing capabilities than the person can provide.

Individual differences in mental competence clearly exist. Are they due to individual differences in basic information processing capacity, in the ability to use general and specific strategies, or in the ability to recognize when problem solving methods are applicable? The answer is "all of the above," but to different degrees in different situations.

Individual Differences in Elementary Processes

There are substantial individual differences in basic information processing capacities. The elementary abilities required to understand language are of particular interest, because language (written and spoken) is so important for human thought. After all, when A wants B to solve a problem, A normally tells B about it.

Language comprehension begins with lexical identification, the recognition that a particular letter or phoneme string is a term in the language. One of the tasks that has been used to study lexical identification is to show people a string of letters, and ask them whether or not it is a word. For example, the string CARD is a word and the string CARG is not. Only common words are used in these studies to avoid testing vocabulary. Even amongst college students, all of whom are good readers by general population standards, there are substantial variations in the speed of lexical identification. In one study, there was approximately a 25 percent variation in speed (560 to 700 milliseconds) between the faster and the slower lexical decision makers.[3,4]

Speech comprehension requires a short term memory capacity, for the comprehender must be able to keep information in mind, while processing the current speech input.

Otherwise the coherency of a message would be lost. The ability to co-ordinate speech analysis with memory can be isolated by the following procedure. A person is asked to read several sentences one at a time. When cued, the person must recall the last word of each sentence. An example of the procedure is:

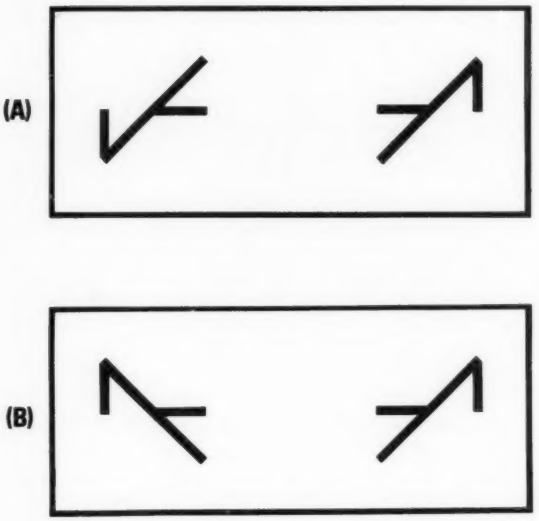
> There are days when the city where I live wakes in the morning with a strange look.
>
> We boys wanted to warn them but we backed down when it came to the pinch.
>
> He stood there at the edge of the crowd while they were singing and he looked bitter.
>
> Recall: look, pinch, bitter.

People vary markedly in the number of sentences that they can process while still being able to recall the words. This is called their reading span. College students show differences of from 2 to 5 sentences. People who show more "verbal aptitude," as measured by tests of reading and paragraph comprehension, seem to have markedly longer spans.[5]

Differences in elementary information processing can also be illustrated in non-verbal problem solving. Figure 1 shows an example of a fairly widely used procedure for evaluating the ability to visualize how objects would look if moved without actually moving them. Two line patterns are shown. The task is to determine whether one pattern can be made identical to the other by rotation in the picture plane. For instance, in Example A in the figure the two patterns can be rotated to match each other. The patterns in Example B cannot be made identical by rotation in the picture plane. It has been found that there is a linear relation between the time required to make a judgement of identity and the size of the angle of rotation. The speed of mental rotation varies markedly across individuals. In a recent study we found that our fastest subject could perform a mental rotation at approximately 2.5 degrees per millisecond, compared to 18.5 degrees for our slowest subject. Men are generally faster at rotation than women, and young adults are, on the average, faster than people in their thirties and beyond.[6] Do individual differences in elementary processes "determine intelligence?" I do not believe this is a sensible question because there is no one thing called intelligence. A more sensible question to ask is whether or not individual differences in elementary processes exert a widespread effect on the mental tasks that people do. The answer seems to be that information processing capacities act as limits, rather in the way that muscle strength limits athletic performance. The correlations between any one measure of information processing ability and a complex mental performance that uses that ability (e.g. the ability to understand stories uses lexical identification) is usually found to be in the .3 range, indicating that about 10 percent of the individual variance in the complex task can be predicted statistically by differences in the elementary task. Utilizing a battery of several basic tasks will increase this correlation, but seldom beyond the .5 range. Such correlations are useful for personnel decision making,[7] but they are certainly not exceptionally high. Much greater differences in basic information processing abilities do appear if one studies extremely disparate groups, e.g. normal and mentally retarded children.[8,9] Pervasive decrements in information processing abilities also appear with advancing age, even during the working years, but their effects may be alleviated or even overridden by an increase in task relevant knowledge.[10] Apparently mental competence is bounded by a person's information processing capacities, but within these bounds very large variations in performance can be achieved by differences in problem solving strategy and by knowledge of a content area.

Figure 1

A mental rotation problem. In (A) the two figures can be rotated "in the mind's eye" to be identical. In (B) they cannot.



Individual Differences in Problem Solving Strategies

Problem solving strategies are plans for executing elementary information processing actions in a coordinated way, so that the solution to a problem is likely to be reached. Strategies can be quite prosaic, such as repeating a person's name upon being introduced in order to increase the likelihood that the name will be recalled later. Other strategies,

such as solving the set of differential equations, are more exotic. Developing a list of the problem solving strategies that might exist would be pointless. There are more strategies than problems, and there are probably individual differences in the use of every conceivable strategy. We can look at examples to illustrate some general principles about individual differences in strategy use.

People vary greatly in how they choose to represent problems to themselves. Figure 2 shows a task that my colleagues and I have used to study people's preferences for visual or verbal representations. It is called the "verification paradigm." We first presented a phrase describing a hypothetical picture. After the participant indicated that the phrase had been understood, a picture was displayed. The task was to decide, as quickly as possible, whether or not the phrase correctly described the picture. Some people memorized the sentence, described the picture to themselves, verbally, and then determined whether or not the phrase and the description were semantically identical. Other people read the phrase and constructed an image of the picture that they expected to see. One strategy puts the burden on verbal information processing, while the other puts the burden on (internally generated) visual image processing.[12,13]

The verification paradigm is of interest as a simple example of the problem of coordinating verbal and visual information. Such coordination is extremely important in human activity. For instance, desert-dwelling Australian aboriginal children use visual strategies to attack memory problems that western-educated children attack by verbal codes.[14] In our work on verification studies we have observed that older people tend to have more difficulty using visual imaging strategies than do young adults, which is consistent with the observation that, on the average, older people have poorer abilities to manipulate visual codes.[15]

Once a problem is coded the code has to be manipulated. Manipulation strategies can vary greatly in their generality, and the use of more general, more powerful strategies is often the characteristic of a good problem solver. A study of strategies for dealing with multiple choice questions on an intelligence test provides a particularly good example.[16] Two strategies were identified. One involved constructing an ideal answer, and then looking for it amongst the multiple choice answers. The other strategy was to search the alternatives provided for what appeared to be a good answer. People who scored high on the test as a whole (by definition, the "more intelligent") used the first strategy more often than did people with low scores. Note that the first strategy is a technique that will work outside of the context of multiple choice tests, while the second one is not.

The acquisition of general problem solving strategies seems to be associated with exposure to formal, western oriented schooling. This does not mean that people don't "learn to think" outside of school; obviously they do. Western schooling does force students to learn how to solve problems stated in abstract and usually verbal terms, outside of their immediate life experiences. As a result, the students must learn to deal with verbal problem statements and to apply general problem solving methods. The two are not exactly the same skills for there are verbal skills that are apart from problem solving. The key to reasoning seems to be developing the idea of a "general problem solving procedure" as an alternative to relying on experience. While the evidence is not conclusive, there are numerous suggestions that the idea of generalized problem solving is a rather sophisticated cultural concept, closely associated with a tradition of literacy. In fact, anthropologists have documented a number of cases in which non-literate respondents simply refused to apply logical reasoning to an abstract problem on the grounds that it was not proper to reason about things outside of one's personal knowledge.[17] Obviously there are variations in personal literacy within the western culture. It has been estimated that 26 million adult Americans are functionally non-literate. Are these people also deficient as general problem solvers?

People may be unable to use a strategy because it places an excessive demand on their information processing capacities. I speculated above that this may be the reason that older people resist using visual imaging strategies. Even more dramatic cases of limitations on the ability to organize information are shown by people who have suffered direct damage to their brains. Some types of damage produce selective loss in capacity for information organization. Memory defects provide an especially interesting case. Patients

Figure 2

The verification paradigm. A phrase is presented. After the observer indicates that the phrase is understood a picture is shown. The observer must decide, as quickly as possible, whether the phrase describes the picture.

PLUS ABOVE STAR

suffering from progressive mental deterioration associated with aging typically perform poorly on general tests of memory, such as those given on intelligence tests. So do patients displaying Korsakoff's syndrome, a possible consequent of alcoholism. Progressive deterioriation patients have great difficulty constructing plans based on information that they obviously once had, e.g. plans for ordering a dinner in a restaurant. Their memory deficit appears to be pervasive and to include reduced access to previously acquired information. Karsakoff syndrome patients can reconstruct such plans. Their memory deficit seems more related to a weakness in organizing new information than in recalling old.[18]

There is another, happier side to the interaction betweeen strategies and information processing capacities. Strategies can be used to "program around" a loss of basic capability. The famous psychologist, B.F. Skinner[19] has offered an interesting case study describing how he has coped with a reduced information processing capacity in his eighties. One of his pieces of advice can be paraphrased as "If something needs to be done, do it now. You may forget." Does reliance on this strategy make Skinner less intelligent, or is he simply being rational?

Most of us would say that Skinner is exceptionally intelligent, because he recognized where a strategy could be used. Intelligent behavior is often limited by people's failure to recognize when strategies are applicable. Like many cognitive phenomena, this can be best illustrated by looking at extreme cases. Mental retardates are notoriously poor at transferring problem solving strategies from one situation to another. On the other hand, the same retardates can be taught quite complex sequences of actions.[8] The principle does not apply just to retardates. An Army officer once described to me the limitations of soldiers who had exceptionally low Armed Forces Qualification Test scores. These soldiers performed tasks that were far more complex than would ever be set for a mental retardate. For instance, they could be trained to perform first level maintainence on tanks. Their commander reported that they had trouble when the task had some element of novelty, such as improvising repairs using materials at hand instead of materials provided specifically for the purpose of the needed repair. Going to yet another situation, a study of students attending the University of Chicago found that the difference between successful and unsuccessful examination takers was often that the successful students could recognize the relevance of acquired information in different contexts, while the unsuccessful students seemed to tie their knowledge more specifically to the situation in which it was learned.[20]

Tests have been developed to evaluate the ability to deal with novel situations; an ability variously called "fluid intelligence"[21] or "dealing with non-entrenched tasks."[22] Studies of the correlations between scores on such tests and scores on the possession of knowledge have shown clearly that the ability to deal with new situations, i.e. the ability to develop new strategies or apply old ones in new situations, is separate from the ability to solve problems by applying set solutions.[23] Since any problem can be treated as a new problem, a talent for dealing with novelty is, by definition, "general intelligence." But how important is it in establishing competence in everyday thinking?

Expertise

A person's ability as a problem solver is highly dependent on his or her experience. We develop expertise in the things we do. The superiority of the content-area expert over the novice includes a superior facility with specific problem-solving methods, an ability to classify problems in a fruitful way, and skill in incorporating new, problem-relevant information into memory.

Studies of the difference between expert and novice problem-solving in physics provide good illustrations of the first two points. In one study[24] professional physicists and beginning students were asked to solve problems in mechanics. The beginning students worked backwards, from the desired solution to the problem. They would consider what the question was, determine what information was required to answer it, and then attack the subproblem of determining the needed information. This is a useful and very general problem-solving procedure. The professional physicists made more use of physics-specific techniques. They would first determine an appropriate physical model for a problem, e.g. a problem involving balances of forces, and then solve a series of equations designed to describe such problems. If they did not discover the correct answer by this method, they considered alternative problem classifications. The procedure of working forward was efficient on easy problems. When faced with a hard problem, the experts could always fall back on the more general method.

A second study[25] isolated the classification process itself. Expert and novice physicists sorted problems into groups of "similar" problems, without actually solving them. The definition of "similar" was left up to the participants. Experts tended to sort problems into classes defined by the physical principles involved: balance of forces, Newton's second law, conservation of momentum, etc. Novices made classifications based upon the surface characteristics of the situation, e.g. problems involving pulleys compared to problems involving sliding blocks. The expert's classifications were tied to problem solving methods more directly than were the novices'.

The next example[26] illustrates the use of expert knowledge to organize newly acquired facts. It was taken from studies of quite a different content area. College students were asked to listen to a "radio broadcast" of an imaginary baseball game. After the broadcast had been heard, the participants were asked to recall the events it described. Those students who were baseball fans could recall much more information than could non-fans. Furthermore, the

advantage was due to the fans' recalling more game-relevant information, rather than recalling more incidental information. It appeared that the fans had a pre-existing idea, or "schema," about how a baseball game should be played. As they listened to the broadcast, they filled out the schema with game-specific information. Thus, when they were asked to recall the game, they were being asked to recall what they perceived as a coherent event. By contrast, the non-fans had to attempt a task that approached recall of isolated bits of information.

These results are quite representative of others that have been obtained in an examination of several fields of expertise, ranging from medicine to computer programming. In addition to telling us about experts, these results make an important general point. Much of the variation in individual differences in mental competence that we observe in everyday life probably springs from individual differences in the sorts of competencies that people have developed, rather than being derived directly from individual differences in information processing capabilities or in general problem-solving ability. Is this intelligence? It is certainly part of mental competence.

Discussion

Whatever happened to intelligence? And what does it all mean for personnel management?

Perhaps the most important message to be obtained from the new approach to intelligence is that human thinking is flexible. Within fairly large bounds, people can trade off strengths and weaknesses in information processing capacity, general problem solving abilities, and expertise in specific content areas. This does not mean that anyone can learn to do anything! There are moderate, but positive correlations between abilities at each of the three levels of mental action.

Minimum levels of information processing capacity are required in order to develop general problem solving skills, and minimum levels of general problem solving skills are needed in order to develop expertise. It is possible to develop tests that evaluate specific aspects of information processing, general reasoning, or knowledge. Such tests can be useful in personnel management. Before deciding what aspect of mental competence is to be measured, careful consideration should be given to the intended use of the results.

Historically, mental testing has been associated with personnel classification. In short-term classification, the purpose of testing is to identify those individuals who will be able to do a well defined task with a minimum of further training. In such situations, the most useful classification information that one can obtain is probably an evaluation of performance on related jobs. This can be obtained either by constructing an appropriate testing situation or by evaluating past experience. Job samples will simultaneously tap the applicants' relevant experience, problem-solving skills, and information-processing capacities. The resulting classification scores should predict performance, but the reason for successful prediction may vary from case to case.

Some personnel classification problems are rather vaguely defined. The purpose may be to predict a multi-attribute, vaguely defined criterion such as "success in the military." Often there is a need to predict behavior in situations where the examinees have not yet developed appropriate expertise, so job sampling is not possible. The selected candidates are usually given substantial further training; so present expertise may not be a crucial issue. Still worse, different candidates may be asked to acquire different types of expertise. All the tester can do is to evaluate those general problem-solving skills, that should predict how candidates will cope with novel environments. It is probably useful to screen out people who do not possess the minimum information-processing capacities required in criterion situations if these are known. (Establishing them may be quite difficult). However, if a person has the minimum information capacities required to succeed, the extent of their success is probably more likely to be due to the possession of problem-solving skills and success in developing expertise than to superior information-processing ability.

The situation changes radically if the purpose of testing is to make the best use of a person, rather than to accept or reject a candidate. This will be called diagnostic assessment. Neuropsychological examinations following brain injury provide an extreme example. Career counselling is a less dramatic diagnostic assessment situation. Here the problem is to obtain a picture of a person without focussing on his or her ability to do a specific task. Diagnostic assessments are typically carried out either because people desire to change their mental working environments or because they can no longer cope with the one they are in. Often diagnostic assessment is required because the person has demonstrated a loss of expertise due either to changes in the environment, such as the introduction of a new technology, or a change in the person such as illness. To illustrate, suppose that a truckdriver can no longer operate a vehicle following a traumatic head injury. The head injury certainly will not have been directed at "driving information," it must have had its affects on basic information capacities used by the high-level action of driving. The problem is to find out what has been affected.

Diagnostic assessment and classification may often be mixed. To continue with the truck driver example, at some point a classification decision would have to be made—should the driver be allowed to drive? Prior to that point the emphasis would be on diagnosis. But at no time would it be useful to decide what the driver's IQ was.

In fact, it may be too bad that we have one word for "intelligence." Individual differences in mental competence do exist, and can be very large. Their assessment is an important part of personnel management. The assessment should depend on a picture, not a number.

References

1. Hunt, E., "On the nature of intelligence,"*Science*, 219 (1983) 141–146.

2. Sternberg, R. J., *The Handbook of Human Intelligence*, Cambridge, Cambridge University Press (1982).

3. Hunt, E., J. E. Davidson, and M. Lansman, "Individual Differences in Long-term Memory Access," *Memory and Cognition*, 9:6, (1981), 599–608.

4. Palmer, J., C. MacLeod, E. Hunt, and J. Davidson, "Information Processing Correlates of Reading," University of Washington Pyschology Department (unpublished paper) (1983).

5. Daneman, M., The Measurement of Reading Comprehension: How Not to Trade Construct Validity for Predictive Power, *Intelligence*, 6 (1983) 331–345.

6. Berg, C., C. Hertzog, and E. B. Hunt, "Age Differences in the Speed of Mental Rotation," *Developmental Psychology*, 18 (1982) 95–107.

7. Allen, M. J. and W. M. Yen, *Introduction to Measurement Theory*, Wadsworth, Inc., Belmont, CA, Brooks/Cole Publishing Co., Monterey, California, (1979).

8. Campione, J., A. L. Brown, and R. A. Ferrara, "Intelligence and Mental Retardation," R. J. Sternberg (Editor), *Handbook of Human Intelligence*, Cambridge, Cambridge University Press (1982).

9. Warren, J., and E. B. Hunt, "Cognitive Processing in Children with Prader-Willi Syndrome," V. A. Holme, S. J. Sulzbacher, and P. L. Pipes (Editors), *Prader-Willi Syndrome*, Baltimore, Maryland, Unviersity Park Press (1981).

10. Hunt, E., and C. Hertzog, "Age Related Changes in Cognition During the Working Years," U. of Washington Psychology Department Technical Report (1981).

11. Salthouse, T. A., *Adult Cognition*, New York, Springer Verlag (1982).

12. MacLeod, C. M., E. B. Hunt, and N. N. Mathews, "Individual Differences in the Verification of Sentence-picture Relationships," *Journal of Verbal Learning and Verbal Behavior* 17, (1978) 493–507.

13. Mathews, H., E. Hunt, and C. MacLeod, "Strategy Choice and Strategy Training in Sentence-picture Verification," *Journal of Verbal Learning and Verbal Behavior*, 19 (1980) 531–548.

14. Kearins, J., "Visual Spatial Memory in Australian Aboriginal Children of Desert Regions," *Cognitive Psycholoby*, 1981, 13, 434.

15. Hunt, E. and J. E. Davidson, "Age Related Changes in Strategies for Sentence Verification," Paper presented at Psychonomics Society meeting (1981).

16. Dillon, R., and R. Stevenson-Hicks, *Effects of Item Difficulty and Method of Test Administration on Eye Scan Patterns During Analogical Reasoning*, Technical Report #1, Contract N660001-80-C0467, Southern Illinois University, (1981).

17. Scribner, S., "Modes of Thinking and Ways of Speaking: Culture and Logic Reconsidered," S. Johnson-Laird, and P. C. Wason, (Editors), *Thinking: Readings in Cognitive Science*, Cambridge, Cambridge University Press (1977).

18. Weingartner, H., J. Gratman, W. Boutelle, W. Kaye, and P. R. Martin, "Forms of Memory Failure. *Science*, 221 (1983) 380–382.

19. Skinner, B. F., "Intellectual Self Management in Old Age, *American Psychologist*, 38 (1983) 239–244.

20. Bloom, G., and L. Broder, *The Problem Solving Processes of College Students*, Chicago, U. of Chicago Press (1950).

21. Cattell, R. B., *Abilities: Their Structure, Growth, and Action*

22. Sternbert, R. J., "Intelligence and Non-entrenchment," *Journal of Education Psychology*, 12, (1981) 10–17.

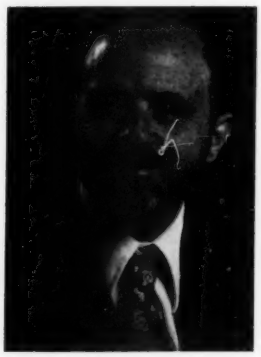
23. Horn, J. L., and G. Donaldson, "Cognitive Development II: Adult Development of Human Abilities," J. Kagan, and O. G. Brim, (Editors), *Consistency and Change in Human Development*, Cambridge, Massachusetts, Harvard University Press (1980).

24. Larkin, J., J. McDermott, D. Simon, and H. Simon, "Expert and Novice Performance in Solving Physics Problems," *Science*, 208 (1980) 1335–1342.

25. Chi, M. T. H., P. J. Feltovitch, and R. Glaser, "Categorization and Representation of Physics Problems by Experts and Novices," *Cognitive Science*, 5 (1981) 121–152.

26. Chiesi, H. L., G. J. Spilich, and J. F. Voss, "Acquisition of Domain-related Information in Relation to High and Low Domain Knowledge," *Journal of Verbal Learning and Verbal Behav-*

Author



Dr. Earl Hunt is a Professor of Psychology at the University of Washington. He has held appointments at Yale, The University of California at Los Angeles, and the University of Sydney in Australia. He has conducted research on cognition (in both people and machines) for the past twenty years. A substantial part of his research deals with the borderline between applied and basic research, focusing on inter- and intra-individual differences in cognitive competence. Professor Hunt is the editor of *Cognitive Psychology*, a leading scientific journal in the field.

Research Notes

Adaptation to New Work Situations

A major characteristic of life in large organizations is that members change positions, roles and locations several times during their tenure. Professor W. H. Mobley at Texas A&M University is supported by the Office of Naval Research to study the transition process and to identify the factors which determine the effectiveness of transitions from one organization to another. The theoretical model comes from earlier ONR supported research in the Marine Corps presented by Mobley in *Employee Turnover: Causes, Consequences, and Control* (Addison-Wesley, 1982). Mobley and his associates are currently applying and testing this theoretical model in an attempt to improve the transition process for Marine Corps personnel in an Okinawa unit rotation.

(Alan W. Lau, ONR)

Creating Long-term Organizational Change

Professor Paul S. Goodman of Carnegie-Mellon University was supported by the Office of Naval Research to evaluate the effects of organizational change on worker satisfaction and productivity and to identify the reasons why some attempts at organizational change have not met with success and others have. A recent book by Goodman entitled *Change in Organizations* (Jossey-Bass, 1982) presents new theoretical insights by some of the top scholars in the organizational field regarding organizational change.

(Alan W. Lau, ONR)

Beyond IQ

Dr. Robert Sternberg of Yale University, who earlier wrote an article for *Naval Research Reviews* on his componential theory of reasoning, will soon publish a book on a more comprehensive theory. This book, titled *Beyond IQ: A Triarchic Theory of Intelligence*, is based largely on research and theory development funded by ONR's Code 442 (Personnel and Training Group). Sternberg believes that the theory goes beyond many previous theories of intelligence in its scope, and answers a broader array of questions about intelligence than has been answered in the past by most single theories. The new theory comprises three subtheories, and may be outlined as follows. The first subtheory concerns the context in which intelligence shows itself and asserts that the way intelligent behavior is defined is much influenced by the socio-cultural context in which the behavior takes place. Behavior is judged intelligent when it is judged to give its owner high potential for adapting to a particular environment; it follows from this that simply translating American intelligence test items into Swahili will not produce a good test for Africans. The second subtheory proposes that even environmentally adaptive behavior may not be considered especially intelligent under all conditions. In particular, account should be taken of whether the behavior involved a problem novel for the person, or whether instead it was an example of a frequently encountered problem for which the person had achieved a fairly routine solution. Within this part of the theory, mental retardates are seen as people who do poorly at developing routine solutions for problems, while the exceptionally gifted are seen as gifted at dealing with novel problems. The third subtheory is essentially the earlier proposed componential theory. It seems likely that Sternberg's discussion of his theory's implications for the field of psychological testing will have a powerful influence on that field in the future. Even more important may be Sternberg's recommendations regarding the design of training to enhance intelligent behavior.

(James T. Lester, ONR)

Cognitive and Instructional Factors in the Acquisition and Maintenance of Skill

The Office of Naval Research is supporting this multifaceted research project of Professor Robert Glaser and his colleagues at the University of Pittsburgh to understand the kind of learning involved in acquiring competence in professions and skilled trades that require hundreds or even thousands of hours of instruction and on-the-job training. Work is focused on specific occupational skills, such as radiological interpretation, physics problem-solving, and route-finding by taxi drivers; on more general job-related capabilities, such as arithmetic problem-solving; and on general relationships between knowledge and competence. Questions being addressed include: (a) What are the qualitative differences in operational performance capability at different stages in the course of training? (b) What do differences in the performance models of experts and novices recommend for theories of learning and instruction? (c) Which aspects of skilled performance depend upon an elaborated understanding of the domain in which skill is exercised, and what ties between conceptual understanding and procedural knowledge are needed for performance to be reliable? (d) What roles does spatial knowledge play in problem-solving expertise? and (e) How do smaller operational units of skill become integrated into robust, flexible and effective procedure structures?

Answers to these questions should enable us to devise instructional strategies and training systems that produce fewer failures of learning or retention, higher levels of skill attainment, more reliable performance, and better transfer to new situations. The explicit goal of the research is to

identify the knowledge structures and information-processing strategies that characterize job expertise and to develop an instructional psychology of long-term learning that is grounded in extensions of current cognitive psychology.

Verbal protocols and anatomical tracings of expert and resident radiologists have revealed that perceptual-recognition processes are critical to effective diagnosis, and that diagnostic flexibility depends on the efficiency of processes involved in building a mental representation of anatomy from the ambiguous film data. New experiments suggest that for classic, textbook cases, there is often a nonlinear learning curve, with intermediate-level trainees showing cognitive overload at the point in their training where they have acquired, but not yet automated, knowledge of the ways in which classic patterns can be disambiguated. Data from interview protocols of students beginning to acquire algebra skills were used in developing a model of comprehending algebra expressions. The model uses a process of parallel activation, enabling simulation of the weakly organized, fragmentary nature of beginning students' knowledge. This investigation extends the application of parallel-activation systems to a domain involving variables and pattern recognition based on structural properties. Experts' and novices' problem-categorization performance and verbal protocols when they plan how to solve physics problems indicate that experts construct representations that include inferences from general principles not mentioned in the problem. In contrast, novices employ representations primarily derived from explicit surface features of the problem. Their difficulties derive from this lack of principled representation of the problem, since it is difficult to reason forward from representations that contain only specific instantiations that are not connected to higher-order abstractions. Hierarchical knowledge structures form the basis of route-finding in both expert and novice taxi drivers, but experts have more detailed knowledge organized into more levels. This hierarchically organized procedural knowledge, rather than better mental maps or greater Euclidean spatial imagery skills, is the source of expert route planning and execution skill.

(Marshall Farr, ONR)

Human Understanding of Complex Sytems

The research described here, which is being done by Dr. Allan Collins, Dedre Gentnew, and Albert Stevens of Bolt Beranek and Newman, Inc. with Office of Naval Research funding seeks to formulate a theory of how humans understand complex causal systems, and how they predict the behavior of such systems under a variety of circumstances. Military personnel who work as boiler technicians, pilots, and electronics technicians all make use of such understanding. The formulation of a theory of this knowledge will make it possible to develop training systems for giving personnel appropriate and powerful conceptions of complex systems, and for teaching them how to choose the conceptions appropriate to particular tasks. To this end, Stevens has been developing a computer system called Steamer, in collaboration with the Naval Personnel Research and Development Center, for automated training on steam-propulsion plants. As a basis for this work, Collins, Gentner and Stevens are seeking to classify and empirically validate different mental models (psychological isomorphs) humans have of physical systems. Such isomorphs represent a person's understanding of the system, and can be qualitative or quantitative, causal or constraint-based, static or dynamic. The nature of these isomorphs is being determined for such domains as steam-propulsion systems, mechanical systems, and electrical systems, using prediction, recall, and reaction-time measures. Further, the research investigates general processing strategies by which models are mapped from one domain to another. This work also makes comparative evaluations of alternative training methods. Current work has emphasized the naive models that people have concerning electricity, fluid flow, and evaporation processes, and has shown how important it is to direct training to students' naive models.

(Marshall Farr, ONR)

☆ U.S. GOVERNMENT PRINTING OFFICE : 1984 O - 437-879

Rear Admiral John B. Mooney, Jr., USN
New Chief of Naval Research

To those who follow the field of oceanography and the achievements of individuals who explore the depths of the world's inner space, Rear Admiral John B. ("Brad") Mooney, Jr., USN, is no stranger.

As the 15th Chief of Naval Research—sworn in October 31, 1983, with additional duties as Chief of Naval Development and Deputy Chief of Naval Material (Technology)—he brings to his new post the tested skills as a naval officer and a noteworthy reputation as an oceanographer. He first gained national recognition as the pilot of the bathyscaph *Trieste II* which found the missing submarine *USS Thresher*, the first U.S. nuclear submarine lost at sea. He located her remains on the floor of the Atlantic at a depth of 8,200 feet.

Two years later he gained recognition for his oceanographic skills, this time on an international scale, when he coordinated the deep submersible search that recovered a lost hydrogen bomb, following an aircraft accident off the coast of Palomares, Spain. He also helped direct the search operations which eventually found the sunken submarine *USS Scorpion*.

Now that he directs the organizations that are primarily responsible for the initial investigation of scientific and engineering exploration for the Navy, Admiral Mooney says he looks forward to the challenge of finding the right balance between conservation and risk. "We must be able to respond to scientific and technical opportunity for the future, and we must be able to extract from technology the solutions to the current problems of the fleet." He adds that he's convinced that there are "unlimited opportunities" to achieve increased capabilities at affordable costs.

It was Admiral Mooney's knowledge of oceanography and his ability to coordinate the advances in ocean engineering devices and techniques that in large part contributed to his earning several military decorations, including three Navy Commendation Medals, three Meritorious Service Medals, and the Legion of Merit. Particularly noteworthy is his designation (in 1973) as a proven subspecialist in oceanography supported by his distinction as the fifth person in U.S. Navy history to qualify as a hydronaut—a deep submergence vehicle operator. (Admiral Mooney also takes special pride in having been awarded the DeMolay Legion of Honor and, as an Eagle Scout from Portsmouth New Hampshire, the Distinguished Eagle Scout Award which he received last year in recognition of his naval career and service to the community.)

During his 30-year career, Admiral Mooney has been primarily a submariner. After graduating from the Naval Academy in 1953 and attending submarine school, his first duty was aboard *USS Burrfish (SSR-312)*, then *USS Sarda (SS-488)*, then back to the submarine school to teach advanced tactics. He served on the staff of the Commander Submarine Force, Atlantic Fleet, and then returned to sea duty as executive officer of the *USS Sea Robin (SS-407)*.

After two years as officer in charge of the *Trieste II* and the Navy Deep Submergence Group, Admiral Mooney commanded the *USS Menhaden (SS-377)* in the Pacific, a tour which included two deployments in the Vietnam combat zone. Under his command, *Menhaden* earned two Battle Efficiency "E" Awards.

He served on the Deep Submergence Staff of the CNO, and in 1971, became the chief staff officer of Submarine Group One. At this time he was recognized for distinguished achievement as the officer in tactical command of a task unit that conducted a three-mile deep recovery in the mid-Pacific.

Since 1973, Admiral Mooney has moved rapidly from one important command to another, assuming more responsibility with each tour of duty: commanding officer of Naval Station, Charleston, South Carolina; deputy director of Deep Submergence Systems Division; commander of the Naval Training Center, Orlando, Florida; director of the Total Force (Manpower, Personnel, and Training) Planning Division in the office of the Chief of Naval Operations; the Oceanographer of the Navy; and now as CNR, CND, and DCNM(T).

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
ARLINGTON, VA. 22217

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

THIRD-CLASS MAIL
POSTAGE & FEES PAID
USN
PERMIT No. GS-9